THE AMERICAN
MATHEMATICAL MONTHLY

AN OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

DEVOTED TO THE INTERESTS OF COLLEGIATE MATHEMATICS

THE AMERICAN MATHEMATICAL MONTHLY, FOUNDED IN 1894 BY BENJAMIN F. FINKEL, WAS PUBLISHED BY HIM UNTIL 1913. FROM 1913 TO 1916 IT WAS OWNED AND PUBLISHED BY REPRESENTATIVES OF FOURTEEN UNIVERSITIES AND COLLEGES IN THE MIDDLE WEST

VOLUME 90

1983

PUBLISHED BY THE ASSOCIATION
MONTPELIER, VT, AND WASHINGTON, DC

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 1 January 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

LETTERS TO THE EDITOR

As an experiment in mathematical communication, the MONTHLY will accept and print Letters to the Editor on *mathematical* subjects of interest to members of the Association.

The emphasis is on mathematics, and, in particular, not on politics of any sort, be it mathematical, academic, national, or global.

The length of a letter should typically be not more than half a printed page.

Letters could be comments on articles (including errors in, additions to, and digressions inspired by them), suggested rephrasings of problems, disagreements with proposed teaching methods, objections to book reviews (based on mathematical grounds), mathematical observations (either new or so old that most of us have forgotten them), possibly overlooked references of possibly wide interest, queries about a mathematical statement (is it known, who did it, where, and when), attractive and challenging "quickie" problems—or anything else that both the readers and the editors of the MONTHLY consider to be of mathematical significance.

Send letters to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, Indiana 47405.

(See page 67 of this issue.)

ANOTHER LOOK AT THE CHEBYSHEV POLYNOMIALS

E. L. ORTIZ
Department of Mathematics, Imperial College of Science and Technology, Queen's Gate, London, England SW7 2BZ

T. J. RIVLIN
Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

The Chebyshev polynomial of degree $n \geqslant 0$ is defined by

$$T_n(x) = \cos n\theta, \quad x = \cos\theta, 0 \leqslant \theta \leqslant \pi. \tag{1}$$

This defines the polynomial on the interval $[-1,1]$ (and hence for all complex numbers). Clearly $T_0(x) = 1$, $T_1(x) = x$, and elementary trigonometric identities show that $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ for $n = 1,2,\ldots$. These polynomials have many interesting properties and have been

---

E. L. Ortiz. I received my doctorate from Buenos Aires University where I was a student of Mischa Cotlar, working in functional analysis. I did postdoctoral research in applied analysis and approximation theory under Cornelius Lanczos at the Institute for Advanced Studies, Dublin, Ireland. I have been professor of mathematics at Buenos Aires University and head of the Numerical Analysis Section of Imperial College, London, England, where I have taught since 1968.

T. J. Rivlin. I received my Ph.D. from Harvard University as a student of J. L. Walsh. After teaching at The Johns Hopkins University and spending a year as a Research Assistant at the Courant Institute for Mathematical Sciences, N.Y.U., I worked as a mathematical analyst in the aircraft industry on Long Island before joining the IBM Research Division in 1959. I have been a Visiting Professor of Computer Science at Stanford, Adjunct Professor of Mathematics at the Graduate Center of the City University of New York, and Science Research Council Senior Research Fellow at the Department of Mathematics, Imperial College, London, where I collaborated with Dr. Ortiz on this piece of work. My major research interest is approximation theory and I am an associate editor of *The Journal of Approximation Theory*. I have written two books: *The Chebyshev Polynomials* (Wiley, 1974) and *An Introduction to the Approximation of Functions* (Dover, 1981).

LETTERS TO THE EDITOR

As an experiment in mathematical communication, the MONTHLY will accept and print Letters to the Editor on *mathematical* subjects of interest to members of the Association.

The emphasis is on mathematics, and, in particular, not on politics of any sort, be it mathematical, academic, national, or global.

The length of a letter should typically be not more than half a printed page.

Letters could be comments on articles (including errors in, additions to, and digressions inspired by them), suggested rephrasings of problems, disagreements with proposed teaching methods, objections to book reviews (based on mathematical grounds), mathematical observations (either new or so old that most of us have forgotten them), possibly overlooked references of possibly wide interest, queries about a mathematical statement (is it known, who did it, where, and when), attractive and challenging "quickie" problems—or anything else that both the readers and the editors of the MONTHLY consider to be of mathematical significance.

Send letters to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, Indiana 47405.

(See page 67 of this issue.)

ANOTHER LOOK AT THE CHEBYSHEV POLYNOMIALS

E. L. ORTIZ
Department of Mathematics, Imperial College of Science and Technology, Queen's Gate, London, England SW7 2BZ

T. J. RIVLIN
Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

The Chebyshev polynomial of degree $n \geqslant 0$ is defined by

$$T_n(x) = \cos n\theta, \quad x = \cos\theta, 0 \leqslant \theta \leqslant \pi. \tag{1}$$

This defines the polynomial on the interval $[-1,1]$ (and hence for all complex numbers). Clearly $T_0(x) = 1$, $T_1(x) = x$, and elementary trigonometric identities show that $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ for $n = 1,2,\ldots$. These polynomials have many interesting properties and have been

---

E. L. Ortiz. I received my doctorate from Buenos Aires University where I was a student of Mischa Cotlar, working in functional analysis. I did postdoctoral research in applied analysis and approximation theory under Cornelius Lanczos at the Institute for Advanced Studies, Dublin, Ireland. I have been professor of mathematics at Buenos Aires University and head of the Numerical Analysis Section of Imperial College, London, England, where I have taught since 1968.

T. J. Rivlin. I received my Ph.D. from Harvard University as a student of J. L. Walsh. After teaching at The Johns Hopkins University and spending a year as a Research Assistant at the Courant Institute for Mathematical Sciences, N.Y.U., I worked as a mathematical analyst in the aircraft industry on Long Island before joining the IBM Research Division in 1959. I have been a Visiting Professor of Computer Science at Stanford, Adjunct Professor of Mathematics at the Graduate Center of the City University of New York, and Science Research Council Senior Research Fellow at the Department of Mathematics, Imperial College, London, where I collaborated with Dr. Ortiz on this piece of work. My major research interest is approximation theory and I am an associate editor of *The Journal of Approximation Theory*. I have written two books: *The Chebyshev Polynomials* (Wiley, 1974) and *An Introduction to the Approximation of Functions* (Dover, 1981).

studied extensively (Cf. Rivlin [**2**]). In view of (1) it is clear that the graph of $y = T_n(x)$ for $-1 \leqslant x \leqslant 1$ lies entirely in the square A: $-1 \leqslant x \leqslant 1$, $-1 \leqslant y \leqslant 1$. Fig. 1 shows the graphs of $y = T_n(x)$, $-1 \leqslant x \leqslant 1$ for $n = 1, 2, \ldots, 30$. Some "white" curves streaking through the square A are clearly visible in Fig. 1. Our primary purpose here is to explain this phenomenon. We wish to suggest that these curves are related to points of intersection of pairs of polynomials, and so we begin by considering such points.

A

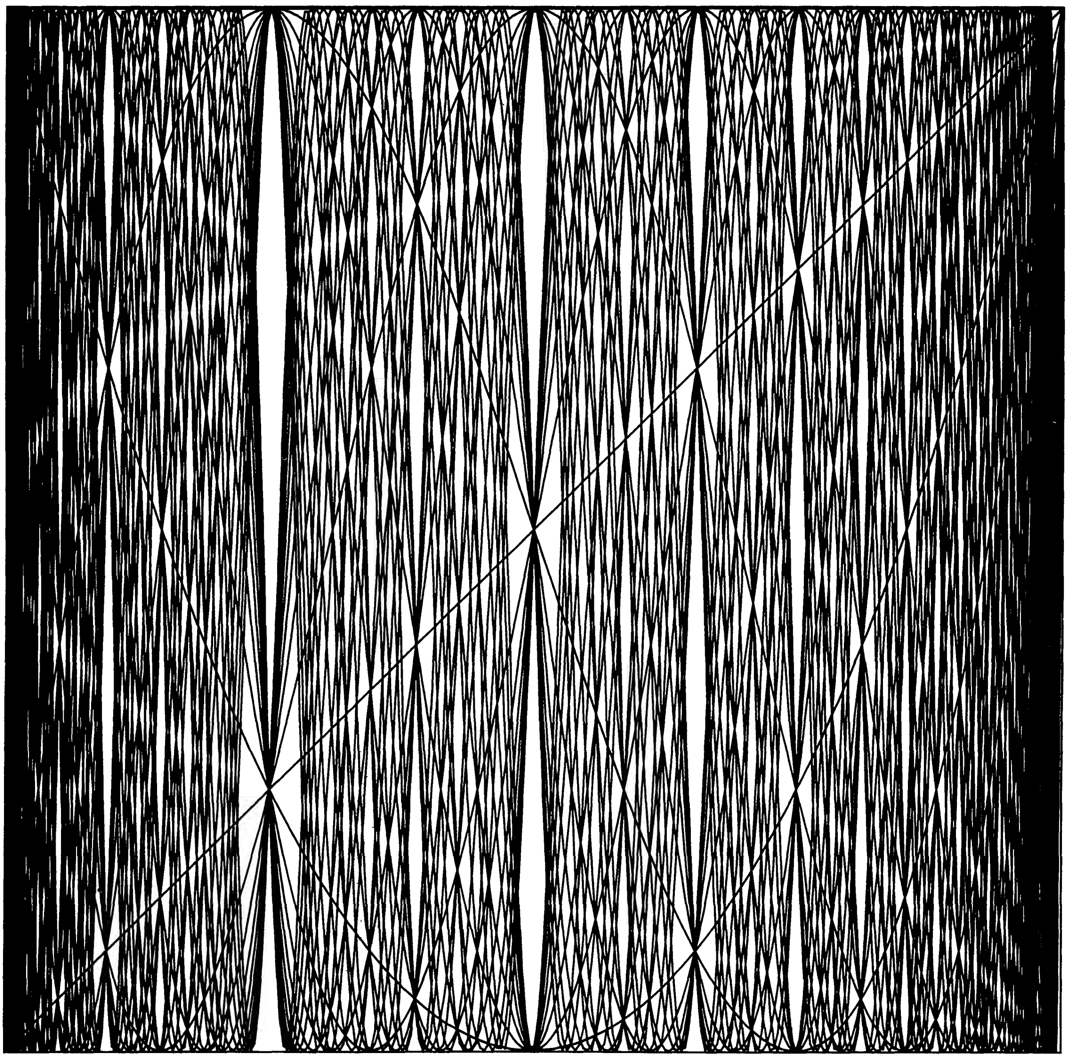


FIG. 1

1. Intersection Points of a Pair of Chebyshev Polynomials. Suppose $1 \leqslant m < n$. $\cos n\theta = \cos m\theta$ if, and only if,

$$\sin \frac{n-m}{2}\theta \cdot \sin \frac{n+m}{2}\theta = 0. \tag{2}$$

If $0 \leqslant \theta \leqslant \pi$, (2) holds only when

$$\theta = \begin{cases} \theta_j = \dfrac{2j\pi}{n+m}, & j = 0, 1, \dots, \left[\dfrac{n+m}{2}\right] \\ \phi_k = \dfrac{2k\pi}{n-m}, & k = 1, \dots, \left[\dfrac{n-m-1}{2}\right], \end{cases}$$

and thus

$$a_j = \cos\frac{2j\pi}{n+m}, \quad j = 0, \dots, \left[\frac{n+m}{2}\right]$$

$$b_k = \cos\frac{2k\pi}{n-m}, \quad k = 1, \dots, \left[\frac{n-m-1}{2}\right]$$

are all n zeros (counting multiplicities) of $T_n(x) - T_m(x) = 0$ in view of (2) and (1). Moreover, we have

$$mT_n'(a_j) + nT_m'(a_j) = 0, \quad -1 < a_j < 1 \tag{3}$$

and

$$mT_n'(b_k) - nT_m'(b_k) = 0, \quad -1 < b_k < 1. \tag{4}$$

To establish (3) and (4) we note that

$$\sin\frac{2j\pi n}{n+m} + \sin\frac{2j\pi m}{n+m} = \sin\frac{2k\pi n}{n-m} - \sin\frac{2k\pi m}{n-m} = 0,$$

and

$$T_n'(x) = n\frac{\sin n\theta}{\sin\theta}, \quad x = \cos\theta.$$

Now suppose that $(x, T_n(x))$ is an intersection point of T_m and T_n in the interior of A. Then $T_n'(x)T_m'(x) \neq 0$. For if $T_n'(x)T_m'(x) = 0$, then (3) and (4) imply that $T_n'(x) = T_m'(x) = 0$ and hence that $T_n(x) = T_m(x) = \pm 1$, contradicting the assumption that $(x, T_n(x))$ is an interior point of A. Thus, in view of (3) and (4)

$$\frac{T_n'(x)}{T_m'(x)} = \begin{cases} -\dfrac{n}{m}, & x = a_j, \\ \dfrac{n}{m}, & x = b_k. \end{cases} \tag{5}$$

(5) provides a basis for our explanation of the existence of the "white" curves. Roughly speaking, for $1 \leqslant m < n$ and $m, n \leqslant 30$, intersection points of type a_j are much more frequent than those of type b_k, and at each intersection point of T_n and T_m of type a_j in the interior of A the slopes of T_n and T_m are of opposite sign and in magnitude in the ratio n to m. We believe that the resulting concatenation of blank areas about these intersection points is what is seen as a "white" curve in Fig. 1.

The connection between the "white" curves and points of intersection of pairs of Chebyshev polynomials receives support from the following considerations.

THEOREM. *If $0 < m \leqslant n$ and $T_m(x) = T_n(x) = y$, then*

$$(1 - T_{n-m}(x))\,(T_2(y) - T_{n-m}(x)) = 0. \tag{6}$$

Proof. The proof follows from the following polynomial identity:

$$T_n^2 - 2T_{n-m}T_nT_m + T_m^2 = 1 - T_{n-m}^2. \tag{7}$$

(The special case $m = 1$ of (7) can be found in Schur [4].) To establish (7) note that for $p \geqslant q$, $2T_pT_q = (T_{p+q} + T_{p-q})$ hence

$$(8) \qquad (T_n - T_m)^2 = (1 - T_{n+m})(1 - T_{n-m})$$

and

$$(9) \qquad 2(1 - T_m T_n) = (1 - T_{n+m}) + (1 - T_{n-m}).$$

(7) now follows upon multiplying (9) by $1 - T_{n-m}$ and subtracting the result from (8). Next put $T_m(x) = T_n(x) = y$ in (7) and the result is (6). (An identity analogous to (7) holds for the Chebyshev polynomials of the second kind.)

Note that $T_{n-m}(x) = 1$ precisely for $x = b_k$, $k = 0, \dots, [(n-m-1)/2]$. Thus the points a_j, $j = 0, 1, \dots, [(n+m)/2]$ all lie on the curve $T_2(y) = T_{n-m}(x)$. This curve is symmetric in the x-axis. It is shown in Fig. 2 for $n - m = q$, $q = 1, 2, 3, 4$. When $q = 2p$ we have $y = \pm T_p(x)$, while for odd q

$$y = \pm\sqrt{\frac{1 + T_q(x)}{2}}\,.$$

$T_2(y) = T_1(x)$ is a parabola, $T_2(y) = T_3(x)$ the folium of Descartes, $x = T_2(t)$, $y = T_3(t)$, $-1 \leqslant t \leqslant 1$.

A

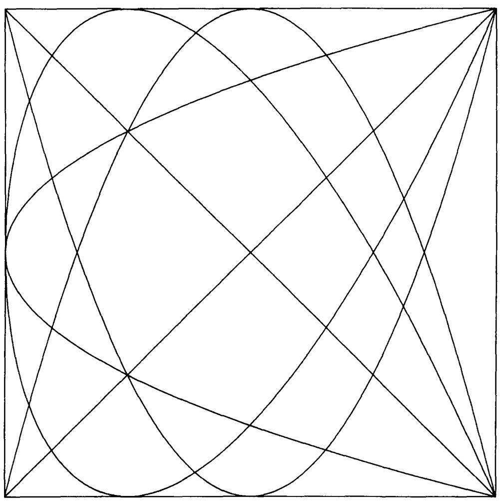


FIG. 2

Observe that the curves for $q = 1, 2, 3, 4$ in Figure 2 are identical to the brightest "white" curves in Fig. 1. Lower values of $n - m = q$ correspond to higher numbers of intersection points of T_n and T_m. Hence for $1 \leqslant m < n \leqslant 30$ we cannot expect to see the curves corresponding to $q > 4$ very clearly in Fig. 1. In the next section we get another view of the "white" curves by applying a suitable homeomorphism to the square A.

2. The Stylized Chebyshev Polynomials. The mapping $S : (x, y) \to (\arccos x, \arccos y) = (x', y')$ is a homeomorphism of the square A onto the square $B : 0 \leqslant x' \leqslant \pi$, $0 \leqslant y' \leqslant \pi$. Let us determine the image of the graph of $T_n(x)$ under S.

$$S : (x, T_n(x)) \to (x', \arccos T_n(\cos x')).$$

Thus $y' = V_n(x') = \arccos(\cos nx')$ is the image of $y = T_n(x)$. Suppose

$$\frac{k\pi}{n} \leqslant x' \leqslant \frac{(k+1)\pi}{n}, \quad k = 0, 1, \dots, n-1.$$

i) If k is even, $\cos nx' = \cos(nx' - k\pi)$, and since $0 \leqslant nx' - k\pi \leqslant \pi$

$$V_n(x') = nx' - k\pi.$$

ii) If k is odd, $\cos nx' = \cos((k+1)\pi - nx')$, and since $0 \leqslant (k+1)\pi - nx' \leqslant \pi$

$$V_n(x') = (k+1)\pi - nx'.$$

In short, then, $y' = V_n(x')$ is a polygonal line issuing from the origin with slope n and with slope changing sign (but maintaining magnitude n) consecutively, at the top and bottom of B, that is at each abscissa $(j\pi)/n$, $j = 1,\ldots, n-1$. (Thus, this image of the Chebyshev polynomial is a perfect spline of degree 1 with knots at $(j\pi)/n, j = 1,\ldots, n-1$. (Cf. Schoenberg [3])).

Each curve $y' = V_n(x')$ passes through (0,0) while each $y = T_n(x)$ is satisfied by (1,1). In order to make the orientation of the polygonal lines agree with those of the Chebyshev polynomials, and to place them on the square A as well, we subject the square B to a linear mapping onto A: $-1 \leqslant x \leqslant 1$, $-1 \leqslant y \leqslant 1$, $L:(x', y') \to (x, y)$ defined by

$$x = 1 - \frac{2}{\pi}x'$$

$$y = 1 - \frac{2}{\pi}y',$$

and thereby obtain as image of $y' = V_n(x')$

$$y = v_n(x) = 1 - \frac{2}{\pi}V_n\left(\frac{\pi}{2}(1-x)\right).$$

We call the piecewise linear curve in A, $y = v_n(x)$, the stylized Chebyshev polynomial of degree n (a name suggested by Kaczmarz and Steinhaus [**1**, p. 43]). Fig. 3 shows $y = v_n(x)$ for $n = 1,\ldots, 5$. Fig. 4 shows the same curves for $n = 1,\ldots, 30$. The "white" curves now seem to be piecewise linear.

A

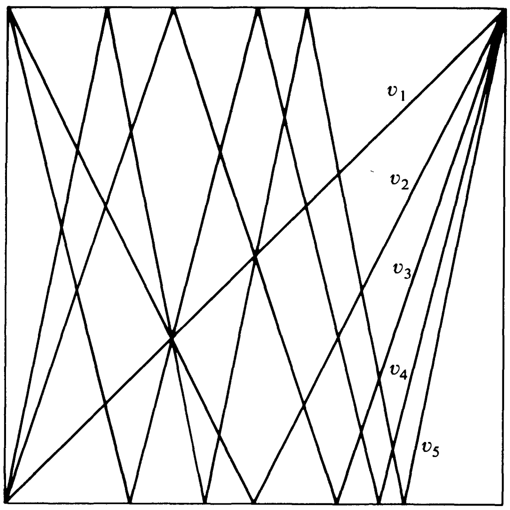


FIG. 3

If $v_n(x) = v_m(x)$ for $0 \leqslant m \leqslant n$, then x is either

$$c_j = 1 - \frac{4j}{m+n}, \quad j = 0,\ldots, \left[\frac{n+m}{2}\right]$$

or

$$d_i = 1 - \frac{4i}{n-m}, \quad i = 1, \ldots, \left[\frac{n-m-1}{2}\right].$$

We claim that at each c_j such that $(c_j, v_n(c_j))$ is an interior point of A the slope of v_n (which is $\pm n$) and the slope of v_m (which is $\pm m$) have opposite signs. For if they had the same signs at c_j and we suppose

$$k < \frac{2jn}{n+m} < k+1$$

and

$$k' < \frac{2jm}{n+m} < k'+1,$$

then k and k' would have the same parity. Thus $k + k' < 2j < k + k' + 2$, which is impossible since $k + k'$ is an even integer. It is this substantial separation of slope (from n to $-m$ or m to

A

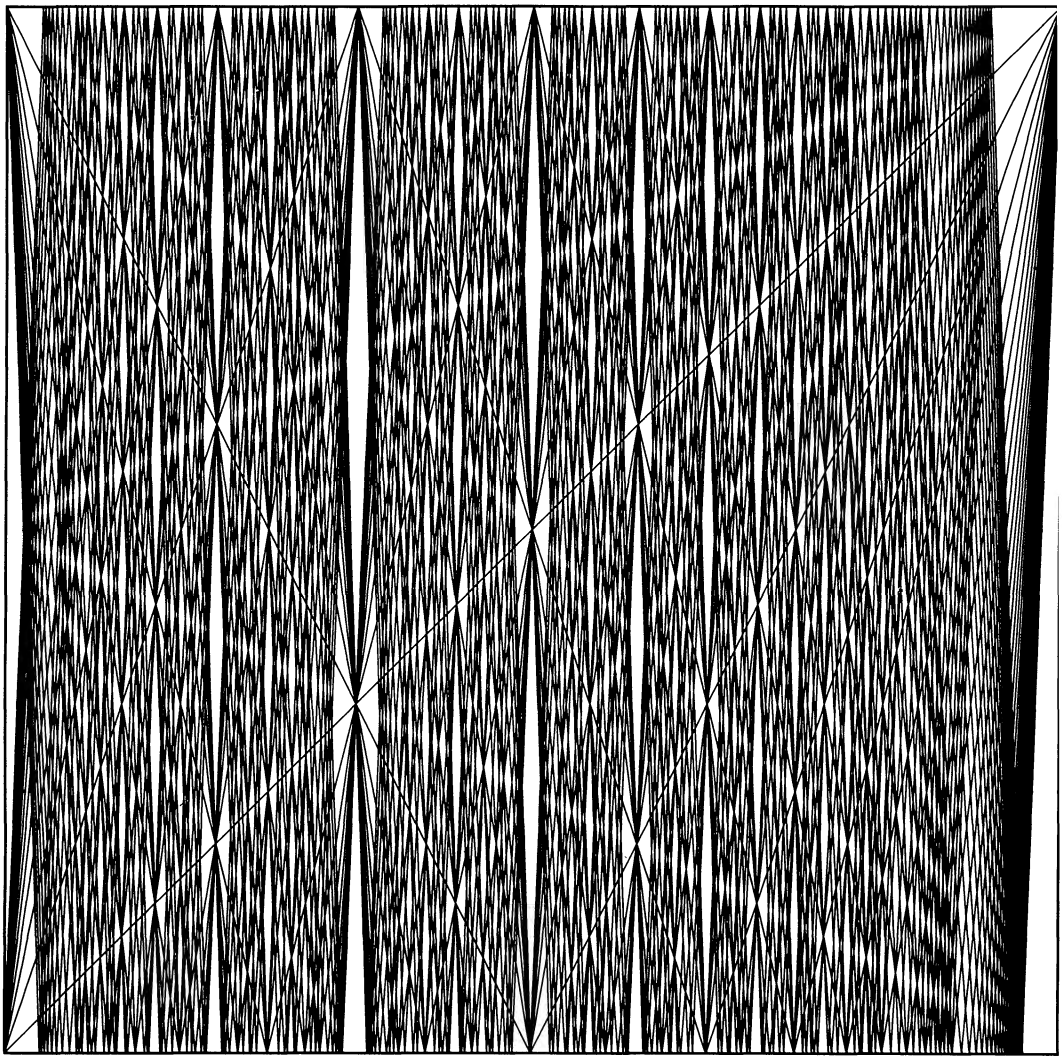


FIG. 4

$-n$) which we believe causes the blank areas that are seen as white lines in Fig. 4, and whose homeomorphic images are the white curves in the Chebyshev case.

The curve $T_2(y) = T_{n-m}(x)$ contains the intersection points $(a_j, T_n(a_j))$ of $y = T_n(x) = T_m(x)$ interior to A, according to the Theorem. Its image under the homeomorphism LS of A onto A is

$$v_2(y) = v_{n-m}(x) \tag{10}$$

which contains the intersection points $(c_j, v_n(c_j))$ of $y = v_n(x) = v_m(x)$ interior to A. (10) is a piecewise linear curve in A which is symmetric with respect to the x-axis. For $y \geq 0$ its equation is

$$y = \frac{1 + v_{n-m}(x)}{2}. \tag{11}$$

The curves (11) are shown in Fig. 5 for $q = n - m = 1,2,3,4$ and their full, symmetric, version (10) is shown in Fig. 6 for $q = n - m = 1,2,3,4$ also. Note that the curves in Fig. 6 fit the brightest "white" curves in Fig. 4 exactly.

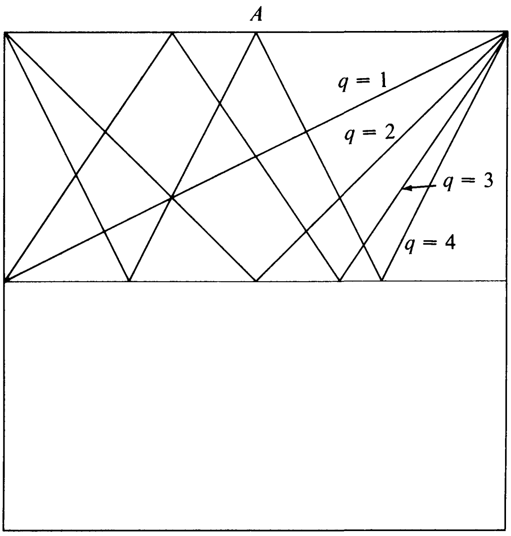


FIG. 5

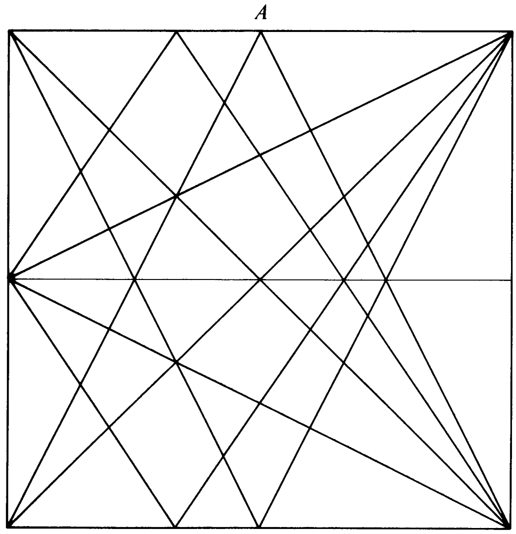


FIG. 6

3. Final Remarks.

(i) *Generality of the Phenomenon.* The phenomenon of the white curves persists for all the Jacobi polynomials, when they are suitably weighted. This is a consequence of an asymptotic formula due to Darboux (see Szegö [4, p. 194]). For example, we have for the Legendre polynomials, as $n \to \infty$

$$\left(\frac{\pi n \sin\theta}{2}\right)^{1/2} P_n(\cos\theta) = \cos\left(\left(n + \frac{1}{2}\right)\theta - \frac{\pi}{4}\right) + O(n^{-1}), 0 < \theta < \pi.$$

The graphs of

$$y = \left(\frac{\pi n}{2}\right)^{1/2}(1 - x^2)^{1/4} P_n(x)$$

for $n = 1, \ldots, 30$ are shown in Fig. 7.

(ii) *Conclusions.* We have attempted to explain the presence of the "white" curves in Fig. 1. We have suggested a mechanism which connects intersection points of pairs of Chebyshev polynomials with the "white" curves based on concatenated blank domains. Strong support is lent to this connection by the fact that the "white" curves coincide with the curves of Fig. 2. Another view of the phenomenon under consideration is offered by the graphs of the stylized Chebyshev poly-

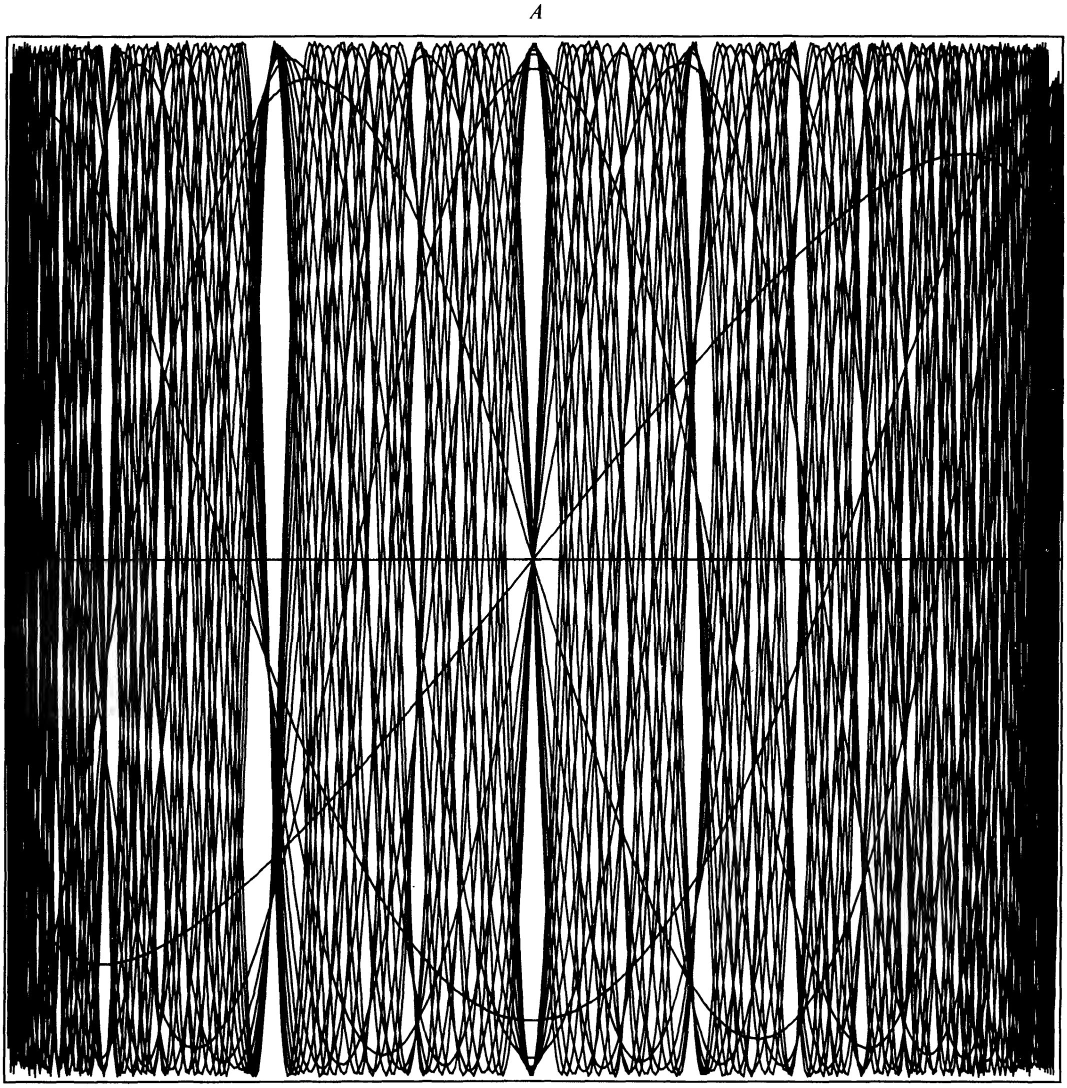


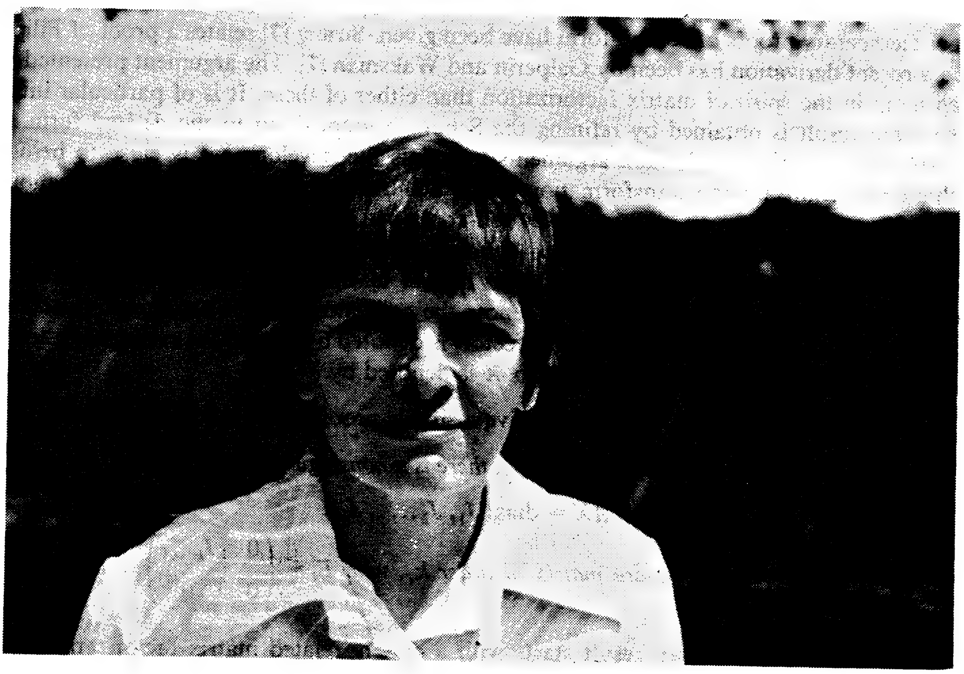
FIG. 7

nomials. The resulting homeomorphic image of Fig. 1 supports the proffered explanation even more strongly, in our opinion.

We wish to thank Alan Konheim and Donald A. Quarles, Jr., of IBM for producing the computer generated figures which appear in this article.

References

1. S. Kaczmarz and H. Steinhaus, Theorie der Orthogonalreihen, Chelsea, New York, 1951.
2. T. Rivlin, The Chebyshev Polynomials, Wiley, New York, 1974.
3. I. J. Schoenberg, Cardinal Spline Interpolation, vol. 12, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1973.
4. I. Schur, Arithmetisches über die Tschebyscheffschen Polynome, Gesammelte Abhandlungen, vol. III, Springer, Berlin, 1973, pp. 422–453.
5. G. Szegö, Orthogonal Polynomials, American Mathematical Society, New York, 1959.



Do you remember Constance Reid, the author of "Hilbert"? The subject above is her sister. Her name is on page 51.

be used to design and administer individual remediation programs. It cannot be over-emphasized that a mathematics appreciation course cannot fulfill its goal if it degenerates into the teaching of arithmetic computations or pre-algebra skills, or if it is limited to a topic such as "consumer mathematics."

6. A large proportion of students enter two-year colleges with little realistic expectation concerning majors. Many of these students have had poor experiences with mathematics and, if there is a general education mathematics requirement which may be satisfied by either a mathematics appreciation course or a pre-calculus/calculus course, they will often elect the mathematics appreciation course. Well into a successful term, the student may begin to think realistically about mathematics requirements of various university majors. Since most majors outside the humanities will necessitate at least some mathematics at a technical level rarely achieved in the typical two-year college mathematics appreciation course, an important service of this course can be to channel these students back into regular sequence mathematics courses. Without violating the spirit of a mathematics appreciation course, it is possible to include a topically organized unit requiring the review and use of elementary algebra and graphing techniques; this may give the student a successful experience in doing mathematics that serve as encouragement to return to regular sequence mathematics courses. (A linear programming unit, for example, requires the students to review or acquire facility with graphing and algebra techniques. Many of the topics suitable for a mathematics appreciation course can be handled in this way.) Students with the experience will frequently place higher in the sequence courses than they would have upon original enrollment, and will go on as solid, though late-blooming, students.

VII. Films. Since students in mathematics appreciation courses frequently have little experience in sustaining interest in regular mathematics lectures, it is usually appropriate in these courses to provide a variety of class activities. Films are a useful but under-utilized medium for mathematics instruction generally. They are especially useful for the mathematics appreciation course.

We list in the Center Section a selection of films about mathematical subjects that are suitable for lay audiences. (Distributor addresses are listed at the end.) Further information on these and other films is available in the booklet *Annotated Bibliography of Films and Videotapes for College Mathematics* by David Schneider (M.A.A., 1980).

VIII. Classroom Aids. Certain topics treated in mathematics appreciation courses are particularly amenable to demonstration with physical or geometric devices. Useful exhibits can often be seen at NCTM meetings. A list of major suppliers of mathematics classroom devices is given in the Center Section.

IX. References. Since many of the topics that arise in mathematics appreciation courses occur nowhere else in the mathematics curriculum, it is quite important that instructors be aware of the expository literature of mathematics that treats its relations to science and society. Student term papers in courses on mathematics appreciation typically tax the instructor's knowledge of the literature more than any other course in the mathematics curriculum.

To aid instructors of mathematics appreciation courses, we list in the Center Section major references that would be suitable for background reading, and as sources for special projects. This list does not include textbooks, partly because we do not wish to endorse some books over others, and partly because texts go in and out of print much more rapidly than the reference classics.

ANSWER TO "PHOTO" ON PAGE 11

Julia Robinson, current president of the American Mathematical Society. Photo by George Bergman.

AN ALGORITHMIC DERIVATION OF THE JORDAN CANONICAL FORM

R. FLETCHER
Department of Mathematics, The University, Dundee, Scotland DD1 4HN

D. C. SORENSEN
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439

1. Introduction. When teaching an introductory course on linear algebra, it might be desirable to introduce Jordan canonical form. However, a standard derivation usually requires more preparation than is justified in such a course. In this note we present an algorithmic derivation of the Jordan canonical form. It requires no preparation other than the Schur decomposition [5] and the solution of linear systems. Because of this the topic can be introduced within the framework of an introductory course with minimal effort.

Other elementary proofs of Jordan form have been given. Strang [7] relates a proof of Filippov [1], and a recent derivation has been by Galperin and Waksman [2]. The argument presented here is much more in the spirit of matrix factorization than either of these. It is of particular interest that the main result is obtained by refining the Schur decomposition to the desired form. The proofs are all based upon induction arguments and thus the development is close to being an algorithmic procedure for the transformation of the Schur decomposition into Jordan canonical form. The derivation given here is completely self-contained and quite simple in our estimation.

We shall use lower case Greek letters to denote scalars, lower case Latin letters to denote column vectors, and upper case Latin letters to denote matrices. All matrices and vectors are assumed to have complex entries. We shall denote the collection of $m \times n$ complex matrices by the symbol $C^{m \times n}$. The transpose of a matrix (vector) is denoted by T, and the conjugate transpose by $*$. The set of eigenvalues of a square matrix A is denoted by $\lambda(A)$.

2. Jordan Canonical Form. Our aim is to give a simple proof of the following theorem.

THEOREM (2.1). *Given any matrix $A \in C^{n \times n}$, there is a nonsingular matrix $X \in C^{n \times n}$, such that*

$$X^{-1}AX = \operatorname{diag}(J_1, J_2, \ldots, J_m)$$

where each $J_i = \lambda_i I + E_i$ is a square matrix of order k_i, with $E_i = \begin{pmatrix} 0 & I_{k_i-1} \\ 0 & 0 \end{pmatrix}$, and $\lambda_i \in C$ for $i = 1, 2, \ldots, m$.

Our proof of this well-known result starts with a closely related matrix factorization due to Schur and systematically transforms this factorization into the Jordan canonical form. This is really a natural approach for anyone who is familiar with the techniques of numerical linear algebra. The Schur decomposition is fundamental to understanding the famous Q-R iterative technique for computing the eigenvalues and eigenvectors of a matrix, and it is evident that the Schur decomposition may be used in almost any situation which might seem to require Jordan form.

R. Fletcher was born on January 29, 1939, in Huddersfield, U.K.; married with two children. MA, Cambridge University, 1960; Ph.D., Leeds University, 1963. Previously Lecturer at Leeds University and Principal Scientific Officer at AERE Harwell; now Reader in Mathematics at the University of Dundee. Research interests: numerical analysis, especially optimization methods and applications, and linear algebra. Leisure interests: bridge, chess, music, squash, swimming, hill walking, jogging.

D. C. Sorensen received his Ph.D. from the Department of Mathematics at the University of California, San Diego, in 1977. He was a member of the Department of Mathematics at the University of Kentucky from 1977-1980 and is now at Argonne National Laboratory in the Mathematics and Computer Science Division. His research interests are primarily in numerical methods for nonlinear optimization.

The proof proceeds in three stages. The Schur decomposition tells us that any matrix $B \in C^{n\times n}$ may be written in the form $B = QRQ^*$ with R upper triangular and Q unitary. The next step is to construct a nonsingular matrix X such that

$$X^{-1}RX = \operatorname{diag}(R_1, R_2, \ldots, R_m)$$

with each R_j being upper triangular and having its main diagonal in the form $\lambda_j I$. The final step is to show how to transform one of the triangular matrices R_j to canonical form. These main steps are described in detail in Lemmas (2.2), (2.8), and (2.12) given below.

Our proof of the Schur decomposition follows the discussion in [**6**, p. 279] very closely but the idea of the proof is essentially the same as the original proof given by Schur in [**5**]. We include it here for the sake of completeness.

LEMMA (2.2) (Schur). *Let $B \in C^{k\times k}$ be any $k \times k$ complex matrix. Then there is a unitary matrix $Q \in C^{k\times k}$ and an upper triangular matrix $R \in C^{k\times k}$ such that*

$$Q^*BQ = R.$$

The diagonal entries of R are the eigenvalues $\lambda(B)$ of B.

Proof. The proof is by induction on k the order of the matrix. The result is clearly true for matrices of order 1. Assume the result for matrices of order k, and suppose that $B \in C^{k+1\times k+1}$. Let λ be an eigenvalue of B, and let q be a corresponding eigenvector such that $q^*q = 1$. Let $U \in C^{k+1\times k}$ be chosen so that the matrix $(q\ U)$ is unitary. Then

$$\begin{pmatrix} q^* \\ U^* \end{pmatrix} B \begin{pmatrix} q & U \end{pmatrix} = \begin{pmatrix} q^*Bq & q^*BU \\ U^*Bq & U^*BU \end{pmatrix} = \begin{pmatrix} \lambda & z^* \\ 0 & B_1 \end{pmatrix}.$$

Now by the induction hypothesis, there is a unitary $Q_1 \in C^{k\times k}$ such that $Q_1^*B_1Q_1 = R_1$ with R_1 upper triangular. Put

$$Q = \begin{pmatrix} q & U \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_1 \end{pmatrix}.$$

Then Q is unitary, and

$$Q^*BQ = R, \text{with } R = \begin{pmatrix} \lambda & r^* \\ 0 & R_1 \end{pmatrix}.$$

This completes the proof. ∎

It is worth noting for the sequel that there is no loss of generality in assuming these eigenvalues occur in a specified order on the diagonal of R.

Our proof of Theorem (2.1) will require knowledge of the existence and uniqueness of solutions to the matrix equation

$$A_1B - BA_2 = S.$$

It is straightforward using the Schur decomposition $Q_j^*A_jQ_j = R_j$ for $j = 1,2$ to transform the above equation to an equivalent equation with upper triangular matrices R_1, R_2 as coefficient matrices in place of A_1, A_2. Thus the following result is completely general.

LEMMA (2.3). *Let R_1 and R_2 be upper triangular matrices in $C^{k_1\times k_1}$ and $C^{k_2\times k_2}$ respectively, and let $S \in C^{k_1\times k_2}$. Then the matrix equation*

$$R_1B - BR_2 = S \tag{2.4}$$

has a unique solution $B \in C^{k_1\times k_2}$ if and only if

$$\lambda(R_1) \cap \lambda(R_2) = \varnothing. \tag{2.5}$$

Proof. The proof is a straightforward induction on k_1 the order of R_1. Let k_2 be any fixed positive integer. The lemma is clearly true for $k_1 = 1$. Assume it is true for upper triangular matrices $\hat{R}_1$ of order less than k_1 and consider the following partitioned form of equation (2.4):

$$(2.6)\qquad \begin{pmatrix} \lambda_1 & r_1^T \\ 0 & \hat{R}_1 \end{pmatrix}\begin{pmatrix} \beta & b^T \\ w & \hat{B} \end{pmatrix} - \begin{pmatrix} \beta & b^T \\ w & \hat{B} \end{pmatrix}\begin{pmatrix} \lambda_2 & r_2^T \\ 0 & \hat{R}_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & s_1^T \\ s_2 & \hat{S} \end{pmatrix}$$

Multiplying the left side of (2.6) out gives the equation

$$(2.7)\qquad \begin{pmatrix} \lambda_1\beta + r_1^T w & \lambda_1 b^T + r_1^T\hat{B} \\ \hat{R}_1 w & \hat{R}_1\hat{B} \end{pmatrix} - \begin{pmatrix} \beta\lambda_2 & \beta r_2^T + b^T\hat{R}_2 \\ w\lambda_2 & w r_2^T + \hat{B}\hat{R}_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & s_1^T \\ s_2 & \hat{S} \end{pmatrix}.$$

The induction hypothesis applied to (2.7) will allow the equations to be solved in the following order

$$\begin{aligned}
&\text{(a)} \quad (\hat{R}_1 - \lambda_2 I)w = s_2 \\
&\text{(b)} \quad (\lambda_1 - \lambda_2)\beta = \sigma_1 - r_1^T w \\
&\text{(c)} \quad \hat{R}_1\hat{B} - \hat{B}\hat{R}_2 = \hat{S} + w r_2^T \\
&\text{(d)} \quad b^T(\lambda_1 I - \hat{R}_2) = s_1^T - r_1^T\hat{B}
\end{aligned}$$

if and only if the matrices R_1 and R_2 satisfy (2.5) This completes the induction. ■

The next lemma refines the Schur decomposition to a form that is very close to the one we seek.

LEMMA (2.8). *Let $R \in C^{n\times n}$ be upper triangular. Then there is a nonsingular $X \in C^{n\times n}$ such that*

$$(2.9)\qquad X^{-1}RX = \operatorname{diag}(R_1, R_2, \ldots, R_m),$$

where

$$R_j = \lambda_j I + U_j, \text{for } j = 1, 2, \ldots, m$$

with each U_j strictly upper triangular, and each λ_j distinct.

Proof. The proof is by induction on n. The result is clearly true when $n = 1$. Assume the result for upper triangular matrices of order less than n. Let $R \in C^{n\times n}$ be upper triangular. The Schur decomposition of a general matrix can be obtained with eigenvalues in any given order. Thus, we may assume without loss of generality that

$$R = \begin{pmatrix} R_1 & S \\ 0 & R_2 \end{pmatrix},$$

where R_1 and R_2 have no eigenvalues in common, and $R_1 = \lambda_1 I + U_1$ with U_1 strictly upper triangular. Now, there is a matrix B of appropriate dimensions which satisfies

$$\begin{pmatrix} I & B \\ 0 & I \end{pmatrix}\begin{pmatrix} R_1 & S \\ 0 & R_2 \end{pmatrix}\begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix},$$

if and only if

$$(2.10)\qquad S = R_1 B - B R_2.$$

The matrix equation (2.10) does have a unique solution B since $\lambda(R_1) \cap \lambda(R_2) = \varnothing$, and the induction hypothesis implies that R_2 may be reduced to block diagonal form as in (2.9). This completes the proof. ■

LEMMA (2.11). *Let $E \in C^{k\times k}$ be of the form*

$$E = \begin{pmatrix} 0 & I_{k-1} \\ 0 & 0 \end{pmatrix}.$$

Then

$$E^TE = \begin{pmatrix} 0 & 0 \\ 0 & I_{k-1} \end{pmatrix},\ E^k = 0,\ Ee_{i+1} = e_i,\ \text{and}\ (I - E^TE)r = \sigma e_1,$$

where e_i is the ith *coordinate vector and $\sigma = e_1^Tr$.*

The proof of Lemma (2.11) is straightforward. The next lemma will show how to reduce the strict upper triangle of the block form in Lemma (2.8) to the form required by Theorem (2.1). In the proof of this lemma we use the notation e_i to denote the ith unit vector without reference to its dimension. In all cases this dimension is meant to be consistent with dimensions implied in context by the matrix partitioning.

LEMMA (2.12). *Let $U \in C^{n\times n}$ be a strictly upper triangular matrix. Then there is a nonsingular matrix X such that*

$$X^{-1}UX = J,$$

where $J =$ diag$(E_1, E_2, \ldots, E_m)$ with each $E_j = \begin{pmatrix} 0 & I_{k_j} \\ 0 & 0 \end{pmatrix}$, and with order $(E_{j+1}) \leqslant$ order$(E_j) = k_j + 1$ for $j = 1, 2, \ldots, m-1$.

Proof. The proof is by induction on n. The result is clearly true for $n = 1$. Assume the result is true for strictly upper triangular matrices U of order $< n$, and let $U \in G^{n\times n}$ be strictly upper triangular. Partition $U = \begin{pmatrix} 0 & u^T \\ 0 & U_1 \end{pmatrix}$. By the induction hypothesis there is a nonsingular X_1 such that

$$X_1^{-1}U_1X_1 = \begin{pmatrix} E_1 & 0 \\ 0 & J_1 \end{pmatrix}, \quad \text{where } J_1 = \text{diag}(E_2, E_3, \ldots, E_m) \tag{2.13}$$

and where order $(E_1) \geqslant$ order (E_j) for $j \geqslant 2$. Then

$$\begin{pmatrix} 1 & 0 \\ 0 & X_1^{-1} \end{pmatrix} U \begin{pmatrix} 1 & 0 \\ 0 & X_1 \end{pmatrix} = \begin{pmatrix} 0 & u^TX_1 \\ 0 & X_1^{-1}U_1X_1 \end{pmatrix}.$$

Partition u^TX_1 as $(u_1^T, u_2^T) = u^TX_1$ consistent with the partitioning in (2.13). With the aid of Lemma (2.11) we obtain

$$\begin{pmatrix} 1 & -u_1^TE_1^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}\begin{pmatrix} 0 & u_1^T & u_2^T \\ 0 & E_1 & 0 \\ 0 & 0 & J_1 \end{pmatrix}\begin{pmatrix} 1 & u_1^TE_1^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} = \begin{pmatrix} 0 & \sigma e_1^T & s^T \\ 0 & E_1 & 0 \\ 0 & 0 & J_1 \end{pmatrix}. \tag{2.14}$$

If $\sigma \neq 0$ then the matrix on the right in (2.14) is similar to

$$\begin{pmatrix} E & e_1s^T \\ 0 & J_1 \end{pmatrix}, \quad \text{with } E = \begin{pmatrix} 0 & e_1^T \\ 0 & E_1 \end{pmatrix}. \tag{2.15}$$

To see this multiply the matrix in (2.14) from the left by diag$(\sigma^{-1}, I, \sigma^{-1}I)$ and from the right by diag$(\sigma, I, \sigma I)$ to obtain (2.15). Note that the order k of E is strictly greater than the order of any diagonal block of J_1 so that $J_1^{k-1} = 0$.

Define $s_i^T = s^TJ_1^{i-1}$ for $i = 1, 2, \ldots, k$. Then

$$\begin{pmatrix} I & e_{i+1}s_i^T \\ 0 & I \end{pmatrix}\begin{pmatrix} E & e_i s_i^T \\ 0 & J_1 \end{pmatrix}\begin{pmatrix} I & -e_{i+1}s_i^T \\ 0 & I \end{pmatrix} = \begin{pmatrix} E & e_{i+1}s_{i+1}^T \\ 0 & J_1 \end{pmatrix}, \ 1 \leqslant i \leqslant k-1.$$

Now, $s_k = 0$ since $J_1^{k-1} = 0$, and it follows that U is similar to the matrix $\begin{pmatrix} E & 0 \\ 0 & J_1 \end{pmatrix}$. On the other hand, if $\sigma = 0$ in (2.14) then a simple permutation of the rows and colums of the matrix on the right of (2.14) shows that U is similar to

$$\begin{pmatrix} E_1 & 0 & 0 \\ 0 & 0 & s^T \\ 0 & 0 & J_1 \end{pmatrix}.$$

By the induction hypothesis there is a nonsingular X_2 such that

$$X_2^{-1}\begin{pmatrix} 0 & s^T \\ 0 & J_1 \end{pmatrix} X_2 = J_2,$$

Thus, U is similar to the matrix $\begin{pmatrix} E_1 & 0 \\ 0 & J_2 \end{pmatrix}$, where J_2 has the desired block diagonal form. Now a simple (block) permutation of the rows and columns will bring this matrix to the proper form. This completes the induction. ■

Theorem (2.1) follows readily from these results. The Schur decomposition shows that a given square matrix is similar to an upper triangular matrix. Lemma (2.8) implies that this triangular matrix is similar to a block upper triangular matrix of the form (2.9). Lemma (2.12) implies that each triangular block is similar to a matrix of the form in Theorem (2.1) because for each diagonal block R_i there is a nonsingular X_i such that

$$X_i^{-1}(\lambda_i I + U_i)X_i = \lambda_i I + \operatorname{diag}(E_1, E_2, \ldots, E_{m_i}).$$

3. Conclusions. The derivation given here is very algorithmic in nature. However, we do not wish to leave the impression that this is a sound numerical computation in any sense. Actual computation of the Jordan canonical form is a formidable task and the reader interested in this computation is referred to [3] for an extensive study of the numerical problem and to [4] for a computer program. Nevertheless, it is interesting to note that the only nonconstructive step in this proof occurs in obtaining the Schur decomposition. For this reason we feel this discussion would fit well in a course concerned with applications or algorithms.

Work supported in part by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under Contract W-31-109-Eng-38. Part of this work was done while the authors were at the University of Kentucky.

References

1. A. F. Filippov, A short proof of the theorem on reduction of a matrix to Jordan form, Vestnik, Moscow University, no. 2 (1971) 18–19.

2. A. Galperin and Z. Waksman, An elementary approach to Jordan theory, this MONTHLY, vol. 87, no. 9 (1981) 728–732.

3. G. H. Golub and J. H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, SIAM Review, vol. 18, no. 4 (1976) 578–619.

4. B. Kagstrom and A. Ruhe, An algorithm for numerical computation of the Jordan normal form of a complex matrix, ACM Transactions on Mathematical Software, vol. 6, no. 3 (1980) 398–419.

5. I. Schur, Uber die characteristchen wurzeln einer linearen substitution mit einer anwendung auf die theorie der integralgleichungen, Math. Ann. vol. 66 (1909) 488–510.

6. G. W. Stewart, Introduction to Matrix Computations, Academic Press, New York, 1973.

7. G. Strang, Linear Algebra and its Applications, Academic Press, New York, 1976.

OF CALCULATIONS PAST AND PRESENT: THE ARCHIMEDEAN ALGORITHM

GEORGE MIEL
Department of Mathematical Sciences, University of Nevada, Las Vegas, NV 89121

There is nothing new in the world except the history you do not know.
Harry S. Truman

1. Introduction. The famous method of inscribed and circumscribed polygons for approximating the constant π, which allowed Archimedes to get the estimates

$$3\frac{1}{7} > \pi > 3\frac{10}{71}, \tag{1.1}$$

was a forerunner of a certain type of algorithms used on today's computers. Stimulated by an article of G. M. Phillips [**49**], which recently appeared in this monthly, we trace the evolution of Archimedes' method, from its geometrical beginning as a means to approximate π to its modern version as an analytical technique for evaluating inverse circular and inverse hyperbolic functions. We show that works of Descartes and Gregory provided crucial stepping stones for the transition from geometry to analysis. The resulting analytical procedure, which was formalized in the nineteenth century by Pfaff and Borchardt, is itself in a class of algorithms dependent on arithmetic and geometric means and on invariant transformations of elliptic integrals. This class of algorithms, investigated in 1971 by B. C. Carlson [**8**], includes Gauss' well-known method for calculating complete elliptic integrals. From the modern standpoint, as shown by (11.1) and (11.4) in the sequel, the Archimedean algorithm can be viewed as a consequence of the duplication theorem applied on a degenerate elliptic integral. In the course of our description of the evolution of Archimedes' method, we establish various historical connections among old and new algorithms. The rich diversity of these connections provides much interesting material for the classroom. For pedagogic reasons, we outline key geometrical arguments in order to display the passage from geometry to analysis, and in the last section, we give a list of pertinent references and computational exercises suited for the classroom.

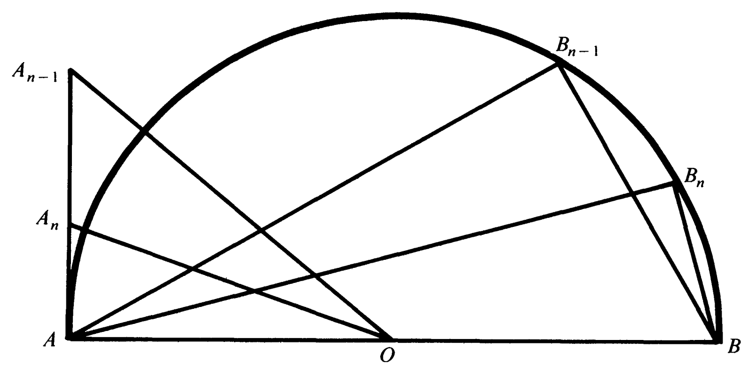


FIG. 1.

George Miel was born in Paris and immigrated to the United States as a teenager. He worked for the Apollo Space Program and for industry abroad. After receiving his Ph.D. at the University of Wyoming in 1976, he held a visiting position for the next two years at the University of Calgary. He then came to the University of Nevada in Las Vegas where he is now Associate Professor. Occasionally he would rather be out scaling wild mountain faces than be in his office writing papers on numerical analysis.

2. An Arithmetical Marvel. Archimedes' book *On the Measurement of a Circle*, written in the 3rd century B.C., consists of three propositions only [**30**], [**31**]. Proposition 3, containing the numerical estimates (1.1), is most interesting to us. We outline and interpret Archimedes' derivation of these estimates. (See Fig. 1.)

Let AA_{n-1} be tangent to a half-circle with center O and diameter AB. Let OA_n bisect the angle AOA_{n-1}. We then have

$$\frac{OA_{n-1}}{OA} = \frac{A_{n-1}A_n}{AA_n}.$$

Consequently,

$$\frac{OA + OA_{n-1}}{OA} = \frac{AA_n + A_{n-1}A_n}{AA_n} = \frac{AA_{n-1}}{AA_n},$$

$$\text{(2.1)} \qquad \frac{OA}{AA_n} = \frac{OA}{AA_{n-1}} + \frac{OA_{n-1}}{AA_{n-1}}.$$

Also,

$$\text{(2.2)} \qquad OA_n^2 = OA^2 + AA_n^2.$$

Letting $\zeta_n = OA / AA_n$ and $\eta_n = OA_n/AA_n$, (2.1) and (2.2) yield the recurrence relations

$$\text{(2.3)} \qquad \zeta_n = \zeta_{n-1} + \eta_{n-1}, \eta_n = \sqrt{\zeta_n^2 + 1}\,.$$

In modern language, ζ_n and η_n are respectively a cotangent and a cosecant and the relations (2.3) represent standard trigonometric identities.

Take B_n on the half-circle so that angle BAB_n equals angle AOA_n. We have

$$2(AA_n) = \frac{AB}{\zeta_n}, BB_n = \frac{AB}{\eta_n}.$$

If angles AOA_n and BAB_n are equal to $\pi/(2^nM)$, for some fixed integer $M \geqslant 3$, then $2(AA_n)$ and BB_n represent sides of regular 2^nM-gons respectively circumscribed and inscribed about a full circle of diameter AB. Taking a unit diameter $AB = 1$, the perimeters of these polygons are given by

$$\text{(2.4)} \qquad P_{2^nM} = \frac{2^nM}{\zeta_n}, \quad p_{2^nM} = \frac{2^nM}{\eta_n}.$$

The two sequences of perimeters given by (2.4) converge monotonically to π.

Archimedes' calculation can be described as follows:

(i) Let $\zeta_0 = \cot \pi/M$ and $\eta_0 = \csc \pi/M$.
(ii) Use (2.3) to find recursively ζ_n and η_n for $n = 1, 2, \ldots, N$.
(iii) Evaluate the perimeters (2.4) with $n = N$.

The final estimates are given by

$$P_{2^NM} > \pi > p_{2^NM}.$$

Archimedes started with circumscribed and inscribed hexagons and he doubled the number of sides four times, namely, he took $M = 6$, $\zeta_0 = \sqrt{3}$, $\eta_0 = 2$, and $N = 4$.

The values of ζ_n and η_n correct to 5 decimal places are listed in Table 1. Due to the backward condition of arithmetic during his time [**30**], Archimedes was unable to get figures to such accuracy. Working with the limited means at his disposal, he controlled the effects of roundoff errors by using his procedure twice, once with consistent underevaluation of the arithmetic operations to get values $\underline{\zeta}_n$ and $\underline{\eta}_n$, and the second time with consistent overevaluation to get

values $\overline{\zeta}_n$ and $\overline{\eta}_n$. The resulting numbers, shown in Table 2 in fractional form and in equivalent decimal notation, satisfy the inequalities

$$\underline{\zeta}_n < \zeta_n < \overline{\zeta}_n, \quad \underline{\eta}_n < \eta_n < \overline{\eta}_n. \tag{2.5}$$

The estimates for π follow from two rounded divisions:

$$\overline{P}_{96} = 3\frac{1}{7} \geqslant \frac{96}{\underline{\zeta}_4}, \quad \underline{p}_{96} = 3\frac{10}{71} \leqslant \frac{96}{\overline{\eta}_4}. \tag{2.6}$$

This systematic use of directed rounding enabled Archimedes to obtain an interval $[\underline{p}_{96}, \overline{P}_{96}]$, guaranteed to contain the exact interval $[p_{96}, P_{96}]$, thus accounting for both the theoretical and the roundoff errors. This strategy was an early example of what is called today interval analysis. This kind of analysis, currently an area of intense research, is described briefly in the next section.

TABLE 1

n	ζ_n	η_n
0	1.73205	2.00000
1	3.73205	3.86370
2	7.59575	7.66130
3	15.25705	15.28979
4	30.54684	30.56320

TABLE 2

n	$\underline{\zeta}_n$	$\underline{\eta}_n$	$\overline{\zeta}_n$	$\overline{\eta}_n$
0	$\frac{265}{153}$ 1.73203	$\frac{306}{153}$ 2.00000	$\frac{1351}{780}$ 1.73205	$\frac{1560}{780}$ 2.00000
1	$\frac{571}{153}$ 3.73203	$\frac{591\frac{1}{8}}{153}$ 3.86356	$\frac{2911}{780}$ 3.73205	$\frac{3013\frac{3}{4}}{780}$ 3.86378
2	$\frac{1162\frac{1}{8}}{153}$ 7.59559	$\frac{1172\frac{1}{8}}{153}$ 7.66095	$\frac{1823}{240}$ 7.59583	$\frac{1838\frac{9}{11}}{240}$ 7.66174
3	$\frac{2334\frac{1}{4}}{153}$ 15.25654	$\frac{2339\frac{1}{4}}{153}$ 15.28922	$\frac{1007}{66}$ 15.25758	$\frac{1009\frac{1}{6}}{66}$ 15.29040
4	$\frac{4673\frac{1}{2}}{153}$ 30.54575		$\frac{2016\frac{1}{6}}{66}$ 30.54798	$\frac{2017\frac{1}{4}}{66}$ 30.56439

In the course of the calculation, Archimedes reduced two fractions to lower terms,

$$\overline{\zeta}_2 = \frac{5924\frac{3}{4}}{780} = \frac{1823}{240}, \qquad \overline{\zeta}_3 = \frac{3661\frac{9}{11}}{240} = \frac{1007}{66}.$$

Without explanation as to how they were obtained, he used the following lower and upper bounds for the square roots involved in the evaluation of $\underline{\eta}_n$ and $\overline{\eta}_n$:

$$\frac{265}{153} < \sqrt{3} < \frac{1351}{780}$$

$$591\frac{1}{8} < \sqrt{349450} \qquad 3013\frac{3}{4} > \sqrt{9082321}$$

$$1172\frac{1}{8} < \sqrt{1373943\frac{9}{16}} \qquad 1838\frac{9}{11} > \sqrt{3380929}$$

$$2339\frac{1}{4} < \sqrt{5472132\frac{1}{16}} \qquad 1009\frac{1}{6} > \sqrt{1018405}$$

$$2017\frac{1}{4} > \sqrt{4069284\frac{1}{36}}$$

There has been much speculation as to the method he employed in the determination of these bounds [**30**, Chapter 4], [**31**, pp. 51–52]. The final estimates (2.6) for π involve the inequalities

$$3\frac{1}{7} > \frac{96 \times 153}{4673\frac{1}{2}}, \quad 3\frac{10}{71} < \frac{96 \times 66}{2017\frac{1}{4}}.$$

Considering that the figures are accurate to about four digits, and that they were obtained without positional arithmetic and without analytic trigonometry, one cannot help wonder at the power displayed by Archimedes' calculation.

3. A Precursor of Interval Analysis. Computational errors range from uncertainties in data due to experimental observation, theoretical errors in approximation techniques, and roundoff errors due to the finite precision of computer arithmetic. Interval analysis provides a means of taking into account all these errors by producing closed intervals of numbers in which exact results are known to lie. The mathematical foundations of interval analysis, which in essence replace point-to-point mappings by set-to-set mappings and approximate equalities by set inclusions, were formulated in 1966 by R.E. Moore [**44**].

Arithmetic operations on closed intervals are defined by

$$[a, b] * [c, d] = \{x * y | a \leqslant x \leqslant b, c \leqslant y \leqslant d\},$$

where $*$ denotes any one of the four operations, with the restriction that $0 \notin [c, d]$ in the case of division. Although interval addition and multiplication are both associative and commutative, the distributive law does not hold in general. During actual computations, exact interval arithmetic is approximated by requiring that the left endpoint be rounded down to the closest machine-representable number and that the right endpoint be similarly rounded up. The representation in a computer of an interval $I = [a, b]$ is $\hat{I} = [\underline{a}, \bar{b}]$, obtained by directed rounding, and if $\hat{*}$ denotes the machine operation corresponding to $*$, then the set inclusions $\hat{I} \supset I$ and $\hat{I}_1 \hat{*} \hat{I}_2 \supset I_1 * I_2$ hold. A computer program executed in rounded-interval arithmetic yields a set of intervals, represented by pairs of machine numbers, each containing the exact result of the corresponding infinite-precision operation.

In interval computations, uncertain input data enter as intervals, error estimates of approximation methods are likewise formulated as intervals and the directed rounding accounts for roundoff errors. The resulting process provides a mathematically rigorous and complete error analysis for computational results. Disadvantages of interval analysis, presently the subject of vigorous research, involve the paucity of supporting hardware and software, the slowness of rounded-interval operations, and the tendency for intervals to grow very large, thus giving overly pessimistic results. An overall view of current research in these areas can be found in Moore [**45**] and in Nickel [**46**].

Archimedes' calculation of π was an early predecessor of interval analysis. The exact intervals I_n, with endpoints defined by (2.4), form a nest of intervals converging to the single point π, thus accounting for the theoretical error inherent in the method. Furthermore, Archimedes accounted for arithmetic errors by systematic directed rounding in order to obtain rounded intervals $\hat{I}_n \supset I_n$. One can roughly simulate his calculations by taking $M = 6$ and computing $\hat{I}_n$ with an arithmetic based on a decimal floating-point representation with a mantissa of four digits. As for Archimedes result (1.1), the final rounded interval, $\hat{I}_4 = [3.141, 3.144]$, provides bounds for π accurate to two decimals. Finally, we note that in a sort of forerunner of Archimedes' calculation of π, Aristarchus (circa 250 B.C.) calculated the equivalent of sines and tangents of certain small angles, also by systematic use of lower and upper bounds, see Heath [**31**, Chapter XII].

4. An Equivalent Formulation. One can show by trigonometry that the second recurrence relation in (2.3) is equivalent to

$$\eta_n = \sqrt{2\zeta_n\eta_{n-1}}\,. \tag{4.1}$$

Using the first relation in (2.3), (4.1), and (2.4) we find that

$$P_{2^nM} = \frac{2P_{2^{n-1}M}p_{2^{n-1}M}}{P_{2^{n-1}M} + p_{2^{n-1}M}} \qquad p_{2^nM} = \sqrt{P_{2^nM}p_{2^{n-1}M}}\,. \tag{4.2}$$

These relations show that the elements of the sequence

$$P_M, p_M, P_{2M}, p_{2M'} P_{4M'} p_{4M'} \ldots$$

can be generated by taking alternatively the harmonic and geometric means of the two preceding elements. The pair of formulas (4.2) is known as the *Archimedean algorithm* [**33**, p. 22].

In modern notation, we have

$$P_M = M\tan\frac{\pi}{M}, \quad p_M = M\sin\frac{\pi}{M}, \tag{4.3}$$

and the essential points of the algorithm depend on the identities

$$\tan\frac{\theta}{2} = \frac{\tan\theta\cdot\sin\theta}{\tan\theta + \sin\theta}, \qquad 2\sin^2\frac{\theta}{2} = \tan\frac{\theta}{2}\cdot\sin\theta,$$

the familiar limits

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = \lim_{\theta\to 0}\frac{\tan\theta}{\theta} = 1,$$

and when $0 < \theta < \pi/2$, the inequalities $\sin\theta < \theta < \tan\theta$ and the monotonicity of $(\sin\theta)/\theta$ and $(\tan\theta)/\theta$.

Divorcing the formulas (4.2) from their geometrical meaning, define

$$a_{n+1} = \frac{2a_nb_n}{a_n + b_n}, \quad b_{n+1} = \sqrt{a_{n+1}b_n}\,, n \geqslant 0. \tag{4.4}$$

It is shown in Section 7 that for any positive initial numbers a_0 and b_0, the sequences a_n and b_n converge to a common limit. Denoting this limit by $A(a_0, b_0)$, the procedure (4.4) defines a nest of intervals converging to the single point $A(a_0, b_0)$.

Archimedes' calculation used recurrence relations different from (4.4) in order to minimize the count of arithmetic operations. However, his method can be described conceptually by

$$A(a_0, b_0) = \pi, \quad a_0 = M\tan\frac{\pi}{M}, \quad b_0 = M\sin\frac{\pi}{M}. \tag{4.5}$$

For completeness, we include here a brief numerical analysis of (4.5).

Numerical stability. Due to roundoff errors, actual computations do not generate the numbers

a_n and b_n, but perturbed numbers $\tilde{a}_n$ and $\tilde{b}_n$ which we represent by

(4.6) $$\tilde{a}_n = a_n(1+\delta), \quad \tilde{b}_n = b_n(1+\delta).$$

The relative error δ is a measure of the number of significant digits in $\tilde{a}_n$ and $\tilde{b}_n$. Suppose that the arithmetic operations in

(4.7) $$\tilde{a}_{n+1} = \frac{2\tilde{a}_n\tilde{b}_n}{\tilde{a}_n + \tilde{b}_n}, \quad \tilde{b}_{n+1} = \sqrt{\tilde{a}_{n+1}\tilde{b}_n}$$

cause negligible roundoff errors. Substitution of (4.6) in (4.7) yields

$$\frac{\tilde{a}_{n+1} - a_{n+1}}{a_{n+1}} = \delta, \quad \frac{\tilde{b}_{n+1} - b_{n+1}}{b_{n+1}} = \delta,$$

thus showing that the relative error is not magnified by the algorithm.

Rate of convergence. Let

$$F(h) = \frac{M}{h}\tan\frac{h\pi}{M} \text{ and } G(h) = \frac{M}{h}\sin\frac{h\pi}{M}.$$

We then have $a_n = F(2^{-n})$, $b_n = G(2^{-n})$, and

(4.8) $$F(h) = \pi + c_1h^2 + c_2h^4 + \ldots + c_mh^{2m} + O(h^{2m+2}),$$

(4.9) $$G(h) = \pi + d_1h^2 + d_2h^4 + \ldots + d_mh^{2m} + O(h^{2m+2}),$$

where c_i and d_i are constants independent of h. Expansions (4.8) and (4.9) with $m = 1$ yield

$$\pi - a_{n+1} = \frac{1}{4}(\pi - a_n) + O(16^{-n-1}),$$

$$\pi - b_{n+1} = \frac{1}{4}(\pi - b_n) + O(16^{-n-1}).$$

The error at the end of each cycle of the iteration is roughly decreased by a factor 4.

Richardson extrapolation. The above rate of convergence makes the Archimedean algorithm much too slow for computation beyond few decimals. However, the expansions (4.8) and (4.9) in even powers of h indicate that the convergence can be accelerated by a method discovered in 1927 by Richardson; see, e.g., [**32**, pp. 239–41]. The procedure is given by

(4.10) $$T_{0n} = a_n, \qquad 0 \leqslant n \leqslant N,$$

(4.11) $$T_{kn} = T_{k-1,n+1} + \frac{T_{k-1,n+1} - T_{k-1,n}}{4^k - 1}, \qquad \begin{array}{l} 1 \leqslant k \leqslant N, \\ 0 \leqslant n \leqslant N-k. \end{array}$$

To show why this procedure speeds up convergence, we proceed as follows. If

$$F_0(h) = F(h),\ F_k(h) = \frac{4^kF_{k-1}(h) - F_{k-1}(2h)}{4^k - 1},$$

then (4.8) implies that

$$F_k(h) = \pi + C_kh^{2k+2} + O(h^{2k+4}),$$

where C_k is a constant independent of h. Since $T_{kn} = F_k(2^{-k-n})$, we get

$$T_{kn} - \pi = 4^{-(k+1)(k+n)}C_k + O(4^{-(k+2)(k+n)}).$$

Thus, the errors in successive elements $T_{k0}, T_{k1}, \ldots, T_{k,N-k}$ of the kth stage of the extrapolation are each time roughly reduced by a factor of $4^{-(k+1)}$. The same conclusion holds if (4.10) is replaced by $T_{0n} = b_n$. It is interesting to note that in this case the elements of the first stage are

given by

$$T_{1n} = b_{n+1} + \frac{1}{3}(b_{n+1} - b_n),$$

which is precisely a formula derived geometrically by Huygens in 1654, see [**33**, p. 30], [47].

Modern computation. Archimedes' calculation corresponds to (4.5) with $M = 6$. When executed on a CDC CYBER 73, a program of (4.4) with $a_0 = 2\sqrt{3}$ and $b_0 = 3$ almost instantaneously printed a result accurate to 12 decimals after 20 iterations. When coupled with the extrapolation (4.11), only 5 iterations were needed to get the same accuracy. In double precision, accuracy to 25 decimals required 42 nonextrapolated iterations compared to only 8 iterations in the extrapolated case.

5. The Method of Equal Perimeters. It is a tribute to the master of antiquity that for almost two millennia, before the discovery of the differential calculus led to a new approach to the problem, the Archimedean method remained the basis for the calculation of π. Of pertinence to this study is a posthumous work of Descartes (1596–1650), in which the polygonal method is regarded from a novel point of view. In a typically Gallic twist, Descartes' approach consisted of doubling the number of sides of regular polygons while keeping the perimeter constant. We derive the method geometrically as follows. (See Fig. 2.)

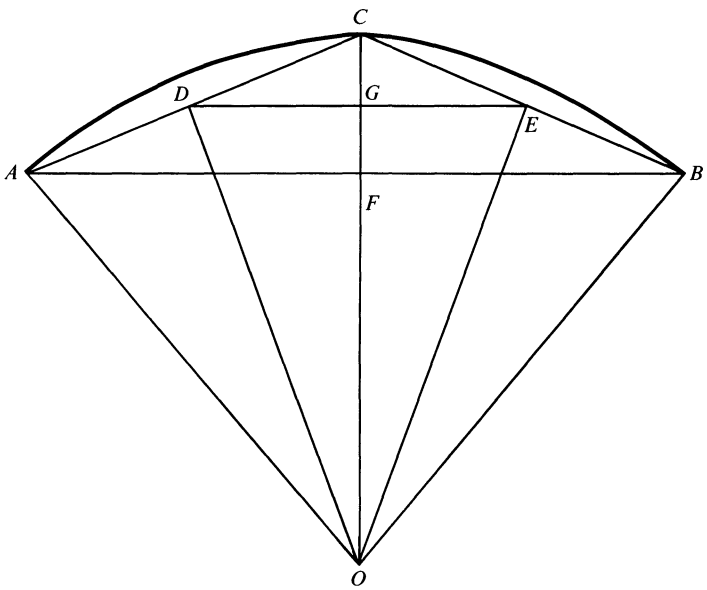


FIG. 2.

Let AB be a side of a $2^{n-1}M$-gon inscribed in a circle of radius OA. Let OC bisect the angle AOB. Draw DE parallel to AB and bisecting CF at G. Since the angle DOE is half the angle AOB, $DE = \frac{1}{2}AB$ is a side of a 2^nM-gon with the same perimeter as the 2^{n-1}M-gon. We have

$$OG = \frac{1}{2}(OF + OC). \tag{5.1}$$

Consideration of the right triangle OCD yields

$$OD = \sqrt{OC \cdot OG}. \tag{5.2}$$

If we denote the radius of the incircle of the 2^nM-gon by r_{2^nM} and the radius of its circumcircle by

R_{2^nM}, then (5.1) and (5.2) yield

$$r_{2^nM} = \frac{1}{2}(r_{2^{n-1}M} + R_{2^{n-1}M}), \quad R_{2^nM} = \sqrt{r_{2^nM}R_{2^{n-1}M}}. \tag{5.3}$$

Thus the elements of the sequence

$$r_M, R_M, r_{2M}, R_{2M}, r_{4M}, R_{4M}, \ldots$$

can be generated by taking alternatively the arithmetic and geometric means of the two preceeding elements. As n increases, these radii tend to the radius of a circle whose circumference equals the common perimeter of the polygons.

Hence, if the initial M-gon has perimeter 2, then r_{2^nM} and R_{2^nM} converge monotonically to $1/\pi$. Observe that in this case, we then have

$$r_M = \frac{1}{M}\cot\frac{\pi}{M}, \quad R_M = \frac{1}{M}\csc\frac{\pi}{M},$$

that comparison with (4.3) yields

$$r_{2^nM} = 1/P_{2^nM}, \quad R_{2^nM} = 1/p_{2^nM}, \tag{5.4}$$

and that (5.4) substituted in (5.3) yields the Archimedean algorithm (4.2). We show in Section 7 that the method of equal perimeters is a special case of an algorithm studied by Pfaff and Borchardt in the nineteenth century.

6. The Quadrature of Conic Sectors. In his first important publication, the *Vera Circuli et Hyperbolae Quadratura*, Gregory (1638–1675) applied the polygonal procedure of Archimedes to find areas of conic sectors. By considering areas rather than perimeters, Gregory was able to deal simultaneously with elliptic and hyperbolic sectors. His approach disclosed the analytical content of the polygonal method and it showed implicitly that the Archimedean algorithm provides a single process for evaluating both circular and hyperbolic arctangents. We interpret Gregory's geometrical argument in modern language. (See Fig. 3.)

Let O be the center of a sector KL of a central conic, let KT and LT be the tangent lines at K and L respectively, and let OT intersect the conic at M. The tangent line at M intersects KT at U

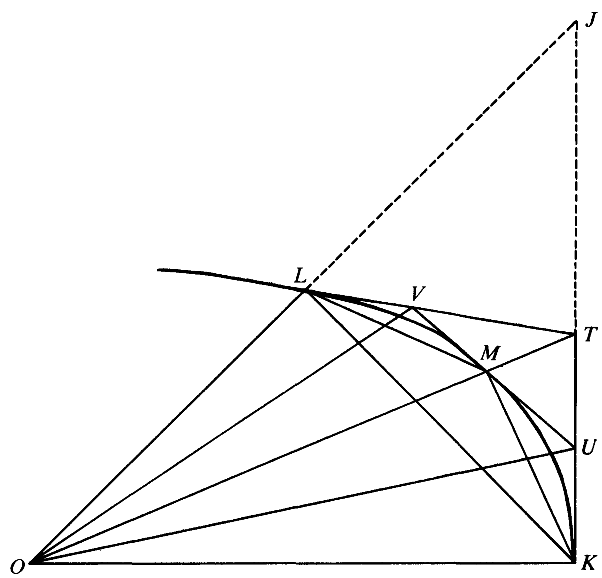


FIG. 3.

and LT at V. The heart of Gregory's strategy is to compare polygonal areas

$$q_1 = OKL, \qquad Q_2 = OKTL,$$
$$q_2 = OKML, \qquad Q_4 = OKUMVL.$$

Basic properties of the central conics enable him to get the relations

$$q_2 = \sqrt{q_1 Q_2}, \quad Q_4 = \frac{2q_2Q_2}{q_2 + Q_2}.$$

Repeating the same construction n times, he gets the relations

$$q_{2^n} = \sqrt{q_{2^{n-1}} Q_{2^n}}, \quad Q_{2^{n+1}} = \frac{2q_{2^n}Q_{2^n}}{q_{2^n} + Q_{2^n}}, \quad n \geqslant 1, \tag{6.1}$$

where q_{2^n} and $Q_{2^{n+1}}$ are polygonal areas composed of 2^n and 2^{n+1} triangles respectively. These polygonal areas converge monotonically to the area of the sector, respectively from below and above in the case of an ellipse, and with directions reversed in the case of a hyperbola.

A minor modification of Gregory's geometry indicates that these polygonal areas can be generated algebraically by the Archimedean algorithm (4.4). Assuming that the angle KOL is acute, extend the lines KT and OL until they meet at J. Let

$$a_0 = OJK, \quad b_0 = q_1 = OKL,$$
$$a_n = Q_{2^n}, \quad b_n = q_{2^n}, \quad n \geqslant 1.$$

It can be shown that $a_1 = 2a_0b_0/(a_0 + b_0)$. We can now change the order of execution of the two relations (6.1), thus getting (4.4).

Consider an ellipse defined parametrically by $x = a\cos t$, $y = b\sin t$. Suppose that the sector KL is given by

$$K\colon (a,0), \quad L\colon (x_1 y_1) = (a\cos t_1, b\sin t_1), \quad 0 < t_1 < \pi/2.$$

One readily deduces that

$$a_n = 2^{n-1}ab\tan\frac{t_1}{2^n}, \quad b_n = 2^{n-1}ab\sin\frac{t_1}{2^n},$$

$$A\left(\frac{a^2y_1}{2x_1}, \frac{ay_1}{2}\right) = \frac{ab}{2}\tan^{-1}\frac{ay_1}{bx_1}. \tag{6.2}$$

For a hyperbolic sector, the same expressions hold with the circular functions replaced by their hyperbolic counterparts. Observe that a special case of Gregory's procedure is in a sense numerically equivalent to Archimedes' calculation of π. Indeed, if a sector consists of one twelfth of a circle of radius $2\sqrt{3}$ then $x_1 = 3$, $y_1 = \sqrt{3}$ and (6.2) then becomes $A(2\sqrt{3}, 3) = \pi$.

We refer to [**16**] and [**62**] for a general review of Gregory's contributions. From the point of view of our study here, there are two points which are particularly pertinent. The *Vera Quadratura* supplied a stepping stone toward the formalization of the Archimedean algorithm as an analytical process for evaluating inverse circular and inverse hyperbolic functions. As shown in the following section, such a formalization was obtained in the nineteenth century by Pfaff and Borchardt by studying sequences closely related to those generated by the Archimedean algorithm. The second point is Gregory's discovery in 1670 of the series,

$$\tan^{-1}x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \ldots, \quad |x| \leqslant 1, \tag{6.3}$$

sometimes mistakenly attributed to Leibniz who found it independently in 1673. This series, combined with certain trigonometric identities, was instrumental in the 1974 computation of π to a million decimals. We deal with the high accuracy computation of π in Section 10.

7. Borchardt's Algorithm. In an unpreserved letter to his teacher, Gauss in 1800 suggested the study of sequences x_n and y_n generated by the relations

$$x_{n+1} = \frac{1}{2}(x_n + y_n), \quad y_{n+1} = \sqrt{x_{n+1}y_n}, \quad n \geqslant 0. \tag{7.1}$$

In his reply [**25**, pp. 234, 284], Pfaff showed that for any positive numbers x_0 and y_0 these sequences converge monotonically to a common limit given by

$$B(x_0, y_0) = \begin{cases} \left(y_0^2 - x_0^2\right)^{1/2}/\cos^{-1}\dfrac{x_0}{y_0}, & 0 \leqslant x_0 < y_0, \\ \left(x_0^2 - y_0^2\right)^{1/2}/\cosh^{-1}\dfrac{x_0}{y_0}, & 0 < y_0 < x_0. \end{cases} \tag{7.2}$$

Pfaff's letter was unpublished in 1881 when Borchardt rediscovered this result which now bears his name [**5**], [**8**], [**61**, Chapter 1].

A quick proof of (7.2) proceeds as follows. Argue by induction that x_n and y_n converge monotonically to the same limit. The ratio $r_n = x_n/y_n$ satisfies $r_{n+1}^2 = \frac{1}{2}(1 + r_n)$. If $x_0 < y_0$, let $\theta = \cos^{-1}r_0$. The quantities $s_n \equiv 2^n\cos^{-1}r_n = \theta$ and $c_n \equiv 4^n(x_n^2 - y_n^2) = x_0^2 - y_0^2$ are independent of n. We have

$$\lim y_n = \lim \frac{2^{-n}|c_n|^{\frac{1}{2}}}{\sin^{-1}\dfrac{2^{-n}|c_n|^{\frac{1}{2}}}{y_n}} = \lim \frac{|c_n|^{1/2}}{s_n} = \frac{\left(y_0^2 - x_0^2\right)^{1/2}}{\theta}.$$

If $y_0 < x_0$, let $\theta = \cosh^{-1}r_0$ and proceed similarly.

We now come to the important observation that *Borchardt's algorithm* (7.1) *is essentially equivalent to the Archimedean algorithm* (4.4). Indeed, if a_n, b_n are generated by (4.4) then $x_n = 1/a_n$, $y_n = 1/b_n$ are generated by (7.1) with

$$A(a_0, b_0) = 1/B(1/a_0, 1/b_0). \tag{7.3}$$

Conversely, if x_n, y_n follow from (7.1), then $a_n = 1/x_n$, $b_n = 1/y_n$ follow from (4.4) with

$$B(x_0, y_0) = 1/A(1/x_0, 1/y_0).$$

Relations (7.2) and (7.3) imply that the common limit of the two sequences generated by the Archimedean algorithm is given by

$$A(a_0, b_0) = \begin{cases} a_0b_0\left(a_0^2 - b_0^2\right)^{-1/2}\cos^{-1}\dfrac{b_0}{a_0}, & a_0 > b_0 \geqslant 0, \\ Ad0b_0\left(b_0^2 - a_0^2\right)^{-1/2}\cosh^{-1}\dfrac{b_0}{a_0}, & b_0 > a_0 > 0. \end{cases} \tag{7.4}$$

The two expressions in (7.4) correspond to (22) and (25) in Phillips [**49**]. The homogeneity properties

$$A(\lambda a_0, \lambda b_0) = \lambda A(a_0, b_0), \quad B(\lambda x_0, \lambda y_0) = \lambda B(x_0, y_0), \quad \lambda > 0, \tag{7.5}$$

are valid.

Observe that Descartes' method of equal perimeters was a predecessor of Borchardt's algorithm with

$$B(r_M, R_M) = \frac{1}{\pi}, \quad r_M = \frac{1}{M}\cot\frac{\pi}{M}, \quad R_M = \frac{1}{M}\csc\frac{\pi}{M}, \tag{7.6}$$

and that the expressions

$$\frac{b}{2}\sqrt{a^2-x^2}\,A\left(\frac{a}{x},1\right)=\frac{ab}{2}\tan^{-1}\frac{\sqrt{a^2-x^2}}{x},\quad 0<x\leqslant a,$$

$$\frac{b}{2}\sqrt{x^2-a^2}\,A\left(\frac{a}{x},1\right)=\frac{ab}{2}\tanh^{-1}\frac{\sqrt{x^2-a^2}}{x},\quad x\geqslant a,$$

describe the algebraic content of Gregory's quadrature of elliptic and hyperbolic sectors.

Relations (7.2), (7.5), and standard identities, imply that

$$\cos^{-1}x=\sqrt{1-x^2}/B(x,1),\, 0\leqslant x\leqslant 1;\ \cosh^{-1}x=\sqrt{x^2-1}/B(x,1),\, x\geqslant 1;$$

$$\sin^{-1}x=x/B\left(\sqrt{1-x^2},1\right),\ -1\leqslant x\leqslant 1;\ \sinh^{-1}x=x/B\left(\sqrt{1+x^2},1\right),\ -\infty<x<\infty;$$

$$\tan^{-1}x=x/B\left(1,\sqrt{1+x^2}\right),\ -\infty<x<\infty;\ \tanh^{-1}x=x/B\left(1,\sqrt{1-x^2}\right),\ -1<x<1;$$

$$\ln x=(x-1)/B\left(\frac{x+1}{2},\sqrt{x}\right),\, x>0.$$

In view of (7.3), these functions can also be evaluated by means of the Archimedean algorithm. The latter, however, is not as computationally efficient as Borchardt's algorithm, since each harmonic mean in (4.4) requires four arithmetic operations compared to two operations for each arithmetic mean in (7.1). In 1972, Carlson [9] coupled Borchardt's algorithm with Richardson extrapolation, in order to obtain a viable technique for the programmed computation of the above functions.

8. One-Sided Sequences. We can uncouple the recurrence relations (7.1) defining Borchardt's algorithm in order to generate each of the two sequences independently of the other. Use the invariant $c=4^n(x_n^2-y_n^2)=x_0^2-y_0^2$ to solve for each of the quantities y_n, x_n in terms of the other. Substitution of the resulting expressions in (7.1) yields

$$x_{n+1}=\frac{x_n}{2}+\sqrt{\left(\frac{x_n}{2}\right)^2-4^{-n-1}c}, \tag{8.1}$$

$$y_{n+1}^2=\frac{y_n^2}{2}\left(1+\sqrt{1+4^{-n}cy_n^{-2}}\right). \tag{8.2}$$

Corresponding relations for the Archimedean algorithm (4.4) are obtained by letting $x_n=1/a_n$, $y_n=1/b_n$:

$$a_{n+1}=\frac{2^{2n+1}}{ca_n}\left(1-\sqrt{1-4^{-n}ca_n^2}\right), \tag{8.3}$$

$$b_{n+1}^2=\frac{2^{2n+1}}{c}\left(\sqrt{1+4^{-n}cb_n^2}-1\right). \tag{8.4}$$

We relate the above recurrence relations to certain old and new algorithms.

The calculation of Archimedes. The two relations (2.3) can easily be uncoupled to get

$$\zeta_{n+1}=\zeta_n+\sqrt{\zeta_n^2+1}, \tag{8.5}$$

$$\eta_{n+1}^2=2\eta_n^2+2\eta_n\sqrt{\eta_n^2-1}. \tag{8.6}$$

Letting $c=-1/M^2$, (8.5) is equivalent to (8.1) with $x_n=\zeta_n/2^nM$ and to (8.3) with $a_n=1/x_n$, whereas (8.6) is equivalent to (8.2) with $y_n=\eta_n/2^nM$ and to (8.4) with $b_n=1/y_n$.

The relation of Bashkara (12th century A.D.). This Indian mathematician [33, p. 22], [4, p. 26] obtained the value 3927/1250 for π by using polygons with 12, 24, 48, 96, 192, and 384 sides. His calculation is based on the formula

$$s_{n+1} = \sqrt{2 - \sqrt{4 - s_n^2}}\,, \tag{8.7}$$

where s_n is the side of a $2^n M$-gon inscribed in a unit circle. The recurrence relation (8.7) is equivalent to (8.4) with $b_n = 2^n M s_n$ and $c = -1/4M^2$.

Viete's infinite product (1593). The first explicit expression for π by an infinite sequence of operations [**33**, p. 26], [**4**, p. 92] can be described by:

$$g_0 = \sqrt{2}\,, g_1 = \sqrt{\frac{1}{2} + \frac{1}{2}\frac{\sqrt{1}}{2}}\,, \quad g_n = \sqrt{\frac{1}{2} + \frac{1}{2}g_{n-1}}\,,$$

$$\lim(g_0 g_1 \cdots g_n) = \frac{4}{\pi}.$$

It turns out that $g_0 g_1 \cdots g_n = y_n$, where y_n is the one-sided sequence corresponding to $B(1, \sqrt{2}) = 1/\sin^{-1}\sqrt{1/2}$.

The relation of Descartes (17th century A.D.). Instead of using the method described in Section 5, Descartes actually generated the one-sided sequence x_n corresponding to (7.6) with $M = 4$. He used an ingenious geometrical construction [**33**, p. **32**], [**4**, p. 118] algebraically equivalent to

$$x_{n+1}(x_{n+1} - x_n) = 4^{-n-3}. \tag{8.8}$$

Solving for positive x_{n+1} in (8.8), we get

$$x_{n+1} = \frac{1}{2}\left(x_n + \sqrt{x_n^2 + 4^{-n-2}}\right),$$

which is the same as (8.1) with $c = -1/16$.

Iterated square root expansions (1961). Thacher [**58**] showed that if $R_1 = \sqrt{2x+2}\,,\ R_{n+1} = \sqrt{R_n + 2}$ then $t_n = 2^n\sqrt{|R_n - 2|}$ converges to $\cos^{-1}x$ if $|x| < 1$ and to $\cosh^{-1}x$ if $|x| \geqslant 1$. For positive x, we have that $t_n = b_{n+1}$, where b_n corresponds to $A(\sqrt{|1 - x^2|}/x, \sqrt{|1 - x^2|})$.

Evaluation of complex logarithms (1981). Carlson's treatment [**9**] of Borchardt's algorithm can be extended to the complex plane. Given a complex number z with $|z| > 1$, consider the iteration

$$\xi_0 = \frac{z+1}{z-1}, \quad \xi_{n+1} = \xi_n + \sqrt{\xi_n^2 - 1}\,, \quad x_n = 2^{-n-1}\xi_n, \tag{8.9}$$

done in complex arithmetic. It turns out that x_n corresponds to

$$B\left(\frac{z+1}{2(z-1)}, \frac{\sqrt{z}}{z-1}\right) = 1/\ln z.$$

Combined with Richardson extrapolation and standard identities, the iteration (8.9) yields viable software for evaluating complex logarithms and other functions [**40**].

9. Carlson's Unified Theory. In 1971, B.C. Carlson [8] presented a unified convergence analysis for all algorithms of the form

$$x_{n+1} = f_i(x_n, y_n), \quad y_{n+1} = f_j(x_n, y_n), \quad n \geqslant 0, \tag{9.1}$$

where

$$f_1(x, y) = \frac{x+y}{2}, \qquad f_2(x, y) = (xy)^{1/2}$$

$$f_3(x, y) = \left(x\frac{x+y}{2}\right)^{1/2}, \quad f_4(x, y) = \left(\frac{x+y}{2}y\right)^{1/2},$$

and where i, j are fixed indices among the numbers $1,2,3,4$. The expressions (9.1) define sixteen algorithms involving arithmetic and geometric means, twelve of which are nontrivial when $i \neq j$, including Borchardt's algorithm when $i = 1, j = 4$. It turns out that when $i \neq j$ the sequences x_n and y_n converge to a common limit of the form

$$L_{ij}(x_0, y_0) = R(a; b, b'; x_0^2, y_0^2)^{-1/2a},$$

where

$$R(a; b, b'; x^2, y^2) = \frac{1}{\beta(a, a')}\int_0^\infty t^{a'-1}(t + x^2)^{-b}(t + y^2)^{-b'}\,dt,$$

$\beta(a, a')$ denotes the beta function, a' is defined by $a + a' = b + b'$, and a, b, b' depend on i and j. We refer to [**8**, pp. 501–3] for the proof. The homogeneity property analogous to (7.5),

$$L_{ij}(\lambda x_0, \lambda y_0) = \lambda L_{ij}(x_0, y_0), \quad \lambda > 0, \tag{9.2}$$

is clearly valid.

For Borchardt's algorithm, the common limit is given by

$$B(x_0, y_0) = L_{14}(x_0, y_0) = R\left(\frac{1}{2}; \frac{1}{2}, 1; x_0^2, y_0^2\right)^{-1}, \tag{9.3}$$

with the R-function in this case degenerating to a function proportional to $\cos^{-1} x_0/y_0$ or $\cosh^{-1} x_0/y_0$. A judicious choice of the starting numbers x_0, y_0 yields an inverse circular, inverse hyperbolic, or logarithmic function. By accelerating the basic iteration with Richardson extrapolation, Carlson [**9**] obtained an efficient and stable method for the computation of these functions.

Certain other algorithms in the class defined by (9.1) deal with fascinating and relevant historical material. The one with $i = 1$ and $j = 3$, discovered by Carlson, has a limit

$$L_{13}(x_0, y_0) = R\left(\frac{1}{4}; \frac{3}{4}, \frac{1}{2}; x_0^2, y_0^2\right)^{-2}, \tag{9.4}$$

which involves an inverse lemniscatic sine,

$$\operatorname{arcsl} r = \int_0^r (1 - t^4)^{-1/2}\,dt, \quad r^2 \leqslant 1.$$

This function is the length of the arc of Bernoulli's lemniscate $r^2 = \cos 2\theta$ from the origin to the point with radial coordinate r. For a description of lemniscate functions, see Marushevich [**36**] or Siegel [**57**]. A particular case of (9.4) is

$$L_{13}(1,0) = 1/A^2 = 16B^2/\pi^2, \tag{9.5}$$

where $A = \operatorname{arcsl} 1$ and $B = \pi/4A$ are the so-called lemniscate constants. A computational history of these constants is given in Todd [**60**]. Whereas Borchardt's algorithm is connected with the rectification of the circle, Carlson's algorithm is related to the rectification of the lemniscate. Moreover, the derivation of (9.4) provides a succinct proof of Fagnano's doubling in 1718 of a lemniscatic arc with ruler and compass. This geometrical construction pointed Euler in 1753 to the addition theorem for general elliptic integrals. This result in turn led Gauss to the development of the theory of lemniscate functions in 1797–98 and to the discovery of general elliptic functions in 1799. For a lively account of the pertinent history, see [**63**], [**53**], [**57**, pp. 1–7].

The best known algorithm of type (9.1) occurs with $i = 1$ and $j = 2$. The corresponding limit $L_{12}(x_0, y_0)$ involves a complete elliptic integral of the first kind,

$$I(x, y) = \int_0^{\pi/2} (x^2\cos^2 t + y^2\sin^2 t)^{-1/2}\,dt. \tag{9.6}$$

Lagrange first recorded the algorithm in 1784–85 by repeatedly using a transformation due to Landen,

$$I(x_{n+1}, y_{n+1}) = I(x_n, y_n);$$

see [**34**, pp. 267, 272], [**63**, pp. 13–16]. Unaware of Lagrange's work, Gauss in 1791, at age fourteen, experimented numerically with the algorithm using $x_0 = 1$ and $y_0 = \sqrt{2}$. He showed in 1799 that the limit in that case is

$$L_{12}(1, \sqrt{2}) = \pi/2A, \tag{9.7}$$

where A is the lemniscate constant [**25**, p. 542]. This result led him from lemniscate functions to general elliptic functions. In 1818, in a work on the motion of planets, he proved the convergence for arbitrary starting values [**24**, pp. 352–355]. The limit, which is given by

$$L_{12}(x_0, y_0) = R\left(\frac{1}{2}; \frac{1}{2}, \frac{1}{2}; x_0^2, y_0^2\right)^{-1} = \pi/2I(x_0, y_0), \tag{9.8}$$

is called the arithmetic-geometric mean of x_0 and y_0.

Of all algorithms defined by (9.1), only the arithmetic-geometric algorithm of Gauss enjoys quadratic convergence, i.e., the error at each step is approximately proportional to the square of the previous error. This means that each cycle of the iteration roughly doubles the number of significant digits. This rapid convergence accounts for the continuing use of the arithmetic-geometric mean in the computation of the elliptic integral $I(x, y)$; see, e.g., [**14**, Algorithms 149, 155]. The next section describes a recent method for computing π based on Gauss's algorithm. The other algorithms in (9.1) converge with a linear rate, the error being roughly reduced by a factor 4 at each step, except for the algorithm with $i = 3$ and $j = 4$, which has a linear rate with factor 2.

Both Gauss's and Borchardt's algorithms are special cases of three-dimensional algorithms investigated by Carlson [**6**], [**7**]. In 1979, Carlson also presented algorithms for the practical computation of incomplete elliptic integrals of all three kinds [**11**]. The iterations, based on successive applications of the duplication theorem, are accelerated by including terms up to fifth order of Taylor series of the integrals. The resulting algorithms, which require only rational operations and square roots, converge linearly with the error roughly decreased by a factor of 4096 in each cycle.

10. Extreme Accuracy Computations. A chronology of the calculation of π can be found in Hobson [**33**], Schepler [**52**], and Wrench [**65**]. The advent of the digital computer allowed computations to hundreds of thousands of digits, and in 1974, the value of π was tabulated to a million decimals [**29**]. Interesting discussions of such extreme computations are given in Beckmann [**4**, Chapters 8, 10] and Davis [15, pp. 55–75].

Borchardt's algorithm, which is a computationally efficient form of the Archimedean algorithm, and the arithmetic-geometric algorithm of Gauss are both in the class of algorithms (9.1) investigated by Carlson. Whereas the former is only linearly convergent with factor 4, the latter is quadratically convergent. Even when accelerated, Borchardt's algorithm in (7.6) is hopelessly too slow for the computation of π beyond few decimals. To date, the most successful techniques for high accuracy computation are based either on Gregory's series (6.3) applied on certain identities expressing π as a linear combination of arctangents, or in a rather surprising twist, on the rapidly converging algorithm of the arithmetic-geometric mean applied on a formula for π recently presented by Salamin [**51**]. We briefly describe these two different techniques.

In 1961, Shanks and Wrench [**56**] used the relation

$$\pi = 24\tan^{-1}\frac{1}{8} + 8\tan^{-1}\frac{1}{57} + 4\tan^{-1}\frac{1}{239}, \tag{10.1}$$

to obtain a value to over 100,200 decimals. The computation, which was verified by using the relation

$$\pi = 48\tan^{-1}\frac{1}{18} + 32\tan^{-1}\frac{1}{57} - 20\tan^{-1}\frac{1}{239}, \tag{10.2}$$

took 8 hours 43 minutes on an IBM 7090. The two relations (10.1) and (10.2) were subsequently used by Guilloud and associates to ultimately calculate π to 1,000,000 decimals [**29**]. The run time on a CDC 7600, not including the verification, was 23 hours 18 minutes, of which 1 hour 7 minutes was used to convert the final result from binary to decimal. Results of statistical tests, which generally support the conjecture that π is normal, are also given in [**29**].

Salamin's identity [**51**], which expresses π in terms of two arithmetic-geometric means, is readily derived as follows. A standard relation due to Legendre can be expressed as

$$I(1,k)J(1,k') + I(1,k')J(1,k) - I(1,k)I(1,k') = \pi/2, \quad k^2 + k'^2 = 1, \tag{10.3}$$

where $I(x,y)$ and $J(x,y)$ are complete elliptic integrals, defined respectively by (9.6) and

$$J(x,y) = \int_0^{\pi/2} \left(x^2\cos^2 t + y^2\sin^2 t\right)^{1/2} dt.$$

By Landen's transformation, if x_n and y_n are generated by (9.1) with $i = 1$ and $j = 2$, then

$$J(x_0, y_0) = \pi\left[x_0^2 - \frac{1}{2}\sum_{n=0}^{\infty}\left(x_n^2 - y_n^2\right)\right]/2L_{12}(x_0, y_0). \tag{10.4}$$

Substitution of (9.8) and (10.4) in (10.3) yields

$$\pi = \frac{4 \cdot L_{12}(1,k)\cdot L_{12}(1,k')}{1 - \sum_{n=1}^{\infty} 2^n\left(z_n^2 + z_n'^2\right)}, \tag{10.5}$$

$$k^2 + k'^2 = 1, \quad z_n^2 = x_n^2 - y_n^2, \quad z_n'^2 = x_n'^2 - y_n'^2,$$

where x_n, y_n correspond to $L_{12}(1,k)$ and x_n', y_n' correspond to $L_{12}(1,k')$.

The computation of π by (10.5) benefits from the rapid convergence of Gauss' algorithm. Since (10.5) defines a continuum of formulas, the computation can be done with one pair k, k' and then verified by using a different pair. The case $k = k'$ causes the two arithmetic-geometric means to coincide, thus halving the work:

$$\pi = \frac{4 \cdot L_{12}\left(1, 1/\sqrt{2}\right)^2}{1 - \sum_{n=1}^{\infty} 2^{n+1} z_n^2}. \tag{10.6}$$

An interesting observation, apparently unreported before, is that the computation of π by (10.6) yields the lemniscatic constant A as a by-product. Indeed, Gauss' result (9.7) and the relation $L_{12}(1, 1/\sqrt{2}) = L_{12}(1,\sqrt{2})/\sqrt{2}$ imply that

$$A = \sqrt{\pi/2D}, \tag{10.7}$$

where D is the denominator in (10.6). Thus, once π is known to the desired accuracy, A can be obtained at the cost of a multiplication, a division, and a square root.

D. E. Knuth and associates at Stanford University have been trying to compute π to 15 million digits, by programming Salamin's identity (10.5) combined with diverse devices, including the Schönhage-Strassen algorithm for multiprecision multiplications [**1**, Section 7.5]. To date, the outcome of this effort has not been recorded in the literature.

11. Classroom Material. In previous sections, we established diverse historical connections and we traced the evolution of the Archimedean algorithm, from its geometrical beginning as a means to approximate π, to its gradual metamorphosis into an analytical method, in the form of Borchardt's algorithm, for calculating certain elementary functions. Carlson showed that Borchardt's algorithm is itself a member of a class of algorithms dependent on invariant transformations of elliptic integrals. This class of algorithms contains the lemniscatic algorithm,

which proves explicitly Fagnano's duplication of a lemniscatic arc by compass and ruler, and the arithmetic-geometric algorithm of Gauss, the only one in the class to enjoy quadratic convergence. Fagnano's construction led Euler successively to the duplication and addition theorems for elliptic integrals. Gauss' algorithm combined with the recent formula of Salamin provides a novel technique for computing π. Moreover, Borchardt's algorithm is a special case of a three-dimensional algorithm based on successive applications of the duplication theorem. We see that the story of the evaluation of π is linked in several ways to invariant transformations of elliptic integrals.

The wealth of connections in this web of old and new algorithms provides considerable pedagogic material. An interesting exercise is to derive the limit (7.4) of the Archimedean algorithm from the duplication theorem, thereby reversing the theme of this article by using a modern result to derive an old one. An induction argument shows that a_n and b_n generated by (4.4) converge to a common limit $\alpha \equiv A(a_0, b_0)$. The degenerate elliptic integral of the first kind,

$$R_c(x,y) = \frac{1}{2}\int_0^\infty (t+x)^{-1/2}(t+y)^{-1}\,dt, \quad x, y > 0, \tag{11.1}$$

clearly has the properties

$$R_c(1,1) = 1, \tag{11.2}$$

$$R_c(\mu x, \mu y) = \mu^{-1/2} R_c(x,y), \quad \mu > 0. \tag{11.3}$$

The duplication theorem [10, p. 279] and (11.3) imply the invariance property

$$R_c(a^{-2}, b^{-2}) = R_c\left(\left[\frac{2ab}{a+b}\right]^{-2}, \left[\frac{2ab}{a+b}\cdot b\right]^{-1}\right) \tag{11.4}$$

Using (11.3) and (11.2), we get

$$R_c\left(a_0^{-2}, b_0^{-2}\right) = R_c\left(a_n^{-2}, b_n^{-2}\right) = R_c\left(\alpha^{-2}, \alpha^{-2}\right) = \alpha$$

Finally, to get (7.4) put $(t + a_0^{-2})/(t + b_0^{-2})$ equal to $\cos^2\theta$ or $\cosh^2\theta$ according as $a_0 > b_0$ or $a_0 < b_0$.

An excellent illustration of the slow rate of convergence of the lemniscatic algorithm compared to the quadratic convergence of the arithmetic-geometric algorithm consists of computing the lemniscatic constant A by (9.5) and (9.7). Applications of the arithmetic-geometric algorithm deal with the computation of the period of a simple pendulum of length L and amplitude ϕ,

$$T = 2\pi\sqrt{\frac{L}{g}}\Big/L_{12}\left(1, \cos\frac{\phi}{2}\right),$$

and the computation of the circumference of an ellipse $x^2/a^2 + y^2/b^2 = 1$,

$$4J(a,b) = 2\pi\left(a^2 - \frac{1}{2}\sum_0^\infty 2^j z_j^2\right)\Big/L_{12}(a,b).$$

Additional applications of elliptic integrals in geometry, mechanics, and potential theory can be found in [10, Section 9.4].

Although not in the class of algorithms defined by (9.1), the method of Heron of Alexandria for finding square roots deserves special consideration [31, pp. 323–6]. Given a positive number a, the method is to make a first guess of $\sqrt{a}$, say x_0, and refine it successively by calculating

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right). \tag{11.5}$$

The iteration can be rewritten in two-dimensional form,

$$x_{n+1} = \frac{1}{2}(x_n + y_n), y_{n+1} = \frac{2}{\frac{1}{x_n} + \frac{1}{y_n}}, \tag{11.6}$$

by letting $y_n = a/x_n$. Since (11.5) happens to be Newton's method [**32**, Chapter 4] applied on the equation $x^2 - a = 0$, we immediately know that x_n converges quadratically to $\sqrt{a}$, and consequently, that y_n also approaches $\sqrt{a}$. The common limit $\sqrt{x_0 y_0}$ of the algorithm (11.6) is sometimes called the arithmetic-harmonic mean of x_0 and y_0. A geometrical diagram of this algorithm is given in Ercolano [**17**]. By using results given in Eves [**18**, p. 36], one can readily get similar diagrams for Archimedes' and Borchardt's algorithms.

A very effective exercise for the classroom is to program the recurrence relations (4.4) with $a_0 = 2\sqrt{3}$ and $b_0 = 3$, thus getting a modern version of Archimedes' calculation of π. The relations (4.10) and (4.11), which are easily incorporated in the program, provide a dramatic illustration of the efficacity of Richardson extrapolation. We described the output of such programs in Section 4. Other numerical results are given in Puritz [**50**], Fox and Hayes [**19**], and Phillips [**49**]. A theoretical exercise is to use the analyticity of $(\tan z)/z$ and $(\sin z)/z$ along the lines of [**40**, p. 747] in order to find bounds for the errors $\pi - T_{kn}$.

The instructor should indicate the difficulties involved in programming the multiprecision arithmetic and the storage allocation needed for the computation of π to a high number of decimals. In single or double precision, however, most techniques involve straightforward programming: the partial sums of Gregory's series (6.3) are easily combined with the formulas (10.1) and (10.2), and Salamin's relation (10.5) is likewise readily programmed. As mentioned earlier, a special case of the latter yields both π and A. A description, well-suited for pedagogic purposes, of a class of formulas for π involving two arctangents is given in [**38**].

It should be mentioned that over 200,000 partial quotients of the simple continued fraction for π have been computed, see Choong et al. [**12**], [**13**] and Gosper [**28**], [**54**]. In this connection, instructive exercises consist of evaluating classical continued fractions, see Olds [**48**, pp. 135–8], by using 2×2 matrices as described in Milne-Thomson [**41**, pp. 109–10] or Frame [**20**].

Other classroom material on the computation of π can be found in Moakes [**42**], [**43**], Goldsmith [**27**], and Maor [**35**]. Buffon's needle experiment is readily simulated on a computer [**4**, p. 164]. The original derivation of Wallis' infinite product makes for an interesting and little-known application of interpolation, see Whiteside [**64**, p. 236ff]. Various classroom notes involving π can be located in a cumulative index for this MONTHLY [**37**, p. 245]. Further pertinent material can be found in Davis [**15**, pp. 52–82], Gardner [**21**], [**22**], [**23**], Anderson [**2**], Goggins [**26**], Nanjundiah [**47**], and Shanks [**55**].

In conclusion, we mention some unsolved problems. Although e^π has been proved to be transcendental, it has not yet been established whether $\pi + e$, πe, or π^e are rational or irrational. While it is known that almost all the irrational numbers must be normal in every base, the normality in the decimal base of such numbers as π, A, $\sqrt{2}$, and e has yet to be proved, though statistical evidence supports the conjecture that they are. Shanks and Wrench [**56**, p. 78] expressed the hope for a theory that would measure the difficulty involved in computing basic mathematical constants; such a theory currently does not exist. Finally, on a more specific level, whereas iterative sequences exist which converge quadratically to π, A, and $\sqrt{2}$, an equally efficient method for computing e is not known, see Baxter [**3**].

Acknowledgment. This article originated while the author was the beneficiary of a stipend as a Faculty Research Participant, Argonne National Laboratory, Division of Education Programs, during the summer 1981.

References

1. A. Aho, J. Hopcroft, and J. Ullman, the Design and Analysis of Algorithms, Addison-Wesley, Reading, Massachusetts, 1974.

2. D. V. Anderson, A polynomial for π, Math. Gaz., 55 (1971) 67–68.
3. L. Baxter, Are π, e, and $\sqrt{2}$ equally difficult to compute?, this MONTHLY, 88 (1981) 165–169.
4. P. Beckmann, A History of π, The Golem Press, Boulder, Colorado, 1977.
5. C. W. Borchardt, Sur deux algorithmes analogues à celui de la moyenne arithmético-géométrique de deux éléments, In Memoriam Dominici Chelini, Collect. Math., L. Cremona, Editor, U. Hoepli, Milan, 1881, pp. 206–212.
6. B. C. Carlson, On computing elliptic integrals and functions, J. Math. and Phys., 44 (1965) 36–51.
7. ______, Hidden symmetries of special functions, SIAM Rev., 12 (1970) 332–345.
8. ______, Algorithms involving arithmetic and geometric means, this MONTHLY, 78 (1971) 496–505.
9. ______, An algorithm for computing logarithms and arctangents, Math. Comp., 26 (1972) 543–549.
10. ______, Special Functions of Applied Mathematics, Academic Press, 1977.
11. ______, Computing elliptic integrals by duplication, Numer. Math., 33 (1979) 1–16.
12. K. Y. Choong, D. E. Daykin, and C. R. Rathbone, Rational approximations to π, Math. Comp., 25 (1971) 387–392.
13. ______, Regular continued fractions for π and γ, UMT 23, Math. Comp., 25 (1971) 403.
14. Collected Algorithms from ACM, ACM Algorithms Distribution Service, IMSL, Inc., 7500 Bellaire Blvd., Houston, TX 77036.
15. P. J. Davis, The Lore of Large Numbers, New Mathematical Library No. 6, Mathematical Association of America, Washington, D.C., 1961.
16. M. Dehn and E. D. Hellinger, Certain mathematical achievements of James Gregory, this MONTHLY, 50 (1943) 149–163.
17. J. Ercolano, A diagram for a square root algorithm, Math. Gaz., 59 (1975) 189–190.
18. H. Eves, A Survey of Geometry, Allyn and Bacon, Boston, 1972.
19. L. Fox and L. Hayes, A further helping of π, Math. Gaz., 59 (1975) 38–40.
20. J. S. Frame, Continued fractions and matrices, this MONTHLY, 56 (1949) 98–103.
21. M. Gardner, New Mathematical Diversions from Scientific American, Simon and Schuster, 1966, pp. 91–102.
22. ______, Mathematical Games, Scientific American (Sept. 1979) 22ff.
23. ______, Mathematical Games, Scientific American (Nov. 1979) 20ff.
24. C. F. Gauss, Werke, vol. 3, Teubner, Leipzig, 1876.
25. ______, Werke, vol. 10, part 1, Teubner, Leipzig, 1917.
26. J. R. Goggins, Formula for $\pi/4$, Math. Gaz., 57 (1973) 134.
27. C. Goldsmith, Calculation of $\ln 2$ and π, Math. Gaz., 55 (1971) 434–436.
28. R. W. Gosper, Jr., Table of the Simple Continued Fraction for π and the Derived Decimal Approximation, Artificial Intelligence Laboratory, Stanford University, Oct. 1975, deposited in UMT file.
29. J. Guilloud and M. Bouyer, Un million de décimales de π, Commissariat à l'Energie Atomique, Paris, 1974.
30. T. L. Heath, The Works of Archimedes, Cambridge University Press, 1897.
31. ______, A History of Greek Mathematics, Oxford University Press, 1921.
32. P. Henrici, Elements of Numerical Analysis, Wiley, New York, 1964.
33. E. W. Hobson, "Squaring the Circle," a History of the Problem, Cambridge, 1913; reprinted by Chelsea, New York, 1953.
34. J. -L. Lagrange, Oeuvres, vol. 2, Gauthier-Villars, Paris, 1868.
35. E. Maor, The history of π on the pocket calculator, J. College Science Teaching, Nov. 1976, 97–99.
36. A. I. Markushevich, The Remarkable Sine Functions, American Elsevier, New York, 1966.
37. K. O. May, Index of the American Mathematical Monthly, vols. 1–80 (1894–1973), The Mathematical Association of America, 1977.
38. G. Miel, An algorithm for the calculation of π, this MONTHLY, 86 (1979) 694–697.
39. ______, Evaluation of complex logarithms and related functions with interval arithmetic, in Interval Mathematics 1980, K. Nickel, Editor, Academic Press, New York, 1980, pp. 407–415.
40. ______, Evaluation of complex logarithms and related functions, SIAM J. Numer. Anal., 18 (1981) 744–750.
41. L. M. Milne-Thomson, The Calculus of Finite Differences, Macmillan, 1960.
42. A. J. Moakes, The calculation of π, Math. Gaz., 54 (1970) 261–264.
43. ______, A further note on machine computation for π, Math. Gaz., 55 (1971) 306–310.
44. R. E. Moore, Interval Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.
45. ______, Methods and Applications of Interval Analysis, Society for Industrial and Applied Mathematics, Philadelphia, 1979.
46. K. Nickel, Editor, Interval Mathematics 1980, Academic Press, New York, 1980.
47. T. S. Nanjundiah, On Huygens' approximation to π, Math. Magazine, 44 (1971) 221–223.

48. C. D. Olds, Continued Fractions, New Mathematical Library No. 9. Mathematical Association of America, Washington, D.C., 1963.
49. G. M. Phillips, Archimedes the numerical analyst, this MONTHLY, 88 (1981) 165–169.
50. C. W. Puritz, An elementary method of calculating π, Math. Gaz., 58 (1974) 102–108.
51. E. Salamin, Computation of π using arithmetic-geometric mean, Math. Comp., 30 (1976) 565–570.
52. H. C. Schepler, The chronology of pi, Math. Magazine, 23 (1950) 165–170, 216–228, 279–283.
53. L. Schlesinger, Über Gauss' Jugendarbeiten zum arithmetisch-geometrischen Mittel, Jber. Deutsch. Math.-Verein., 20 (1911) 396–403.
54. D. Shanks, Review 15, Math. Comp., 31 (1977) 1044.
55. ______, Quartic approximations for π, Abstract No. 80T-A183, Abstracts Amer. Math. Soc., 1 (1980) 558.
56. D. Shanks and J. W. Wrench, Jr., Calculation of π to 100,000 decimals, Math. Comp., 16 (1962) 76–99.
57. C. L. Siegel, Topics in Complex Function Theory, vol. 1, Wiley-Interscience, New York, 1969.
58. H. C. Thacher, Jr., Iterated square root expansions for the inverse cosine and inverse hyperbolic cosine, Math. Comp., 15 (1961) 399–403.
59. J. Todd, A problem on arc tangent relations, this MONTHLY, 56 (1949) 517–528.
60. ______, The lemniscate constants, Commun. ACM, 18 (1975) 14–19, 462.
61. ______, Basic Numerical Mathematics, Academic Press, New York, 1981.
62. H. W. Turnbull, Editor, James Gregory Tercentenary Memorial Volume, published for the Royal Society of Edinburgh, London, 1939.
63. G. N. Watson, The marquis and the land-agent; a tale of the eighteenth century, Math. Gaz., 17 (1933) 5–17.
64. D. T. Whiteside, Patterns of mathematical thought in the latter seventeenth century, Archive for History of Exact Sciences, vol. 1, 1960–1962.
65. J. W. Wrench, Jr., The evolution of extended decimal approximations to π, Math. Teacher, 53 (1960) 644–650.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N IN4.*

DON'T TRY TO SOLVE THESE PROBLEMS!

RICHARD K. GUY
Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

Such an exhortation will likely produce the opposite effect, but I'm serious, and I'll explain why. This article has been in mind for some time, but its eruption is triggered by a proposal from

Schmuel Schreiber, Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel.

Problem 0. For an integer a define the set S_a inductively by

$$(1)\ a \in S_a,\ (2)\ \text{if } k \in S_a, \text{ then } 2k+2 \in S_a,\ (3)\ \text{if } k \in S_a, \text{ then } 3k+3 \in S_a.$$

Equivalently, define a function $s_a(n)$ on the integers by

$$(1) \qquad s_a(1) = a, \quad (2)\ s_a(2k) = 2s_a(k) + 2, \quad (3)\ s_a(2k+1) = 3s_a(k) + 3.$$

For $a < -3$ or $a > 2$ is s_a injective? Or does S_a contain repeated elements?

48. C. D. Olds, Continued Fractions, New Mathematical Library No. 9. Mathematical Association of America, Washington, D.C., 1963.

49. G. M. Phillips, Archimedes the numerical analyst, this MONTHLY, 88 (1981) 165–169.

50. C. W. Puritz, An elementary method of calculating π, Math. Gaz., 58 (1974) 102–108.

51. E. Salamin, Computation of π using arithmetic-geometric mean, Math. Comp., 30 (1976) 565–570.

52. H. C. Schepler, The chronology of pi, Math. Magazine, 23 (1950) 165–170, 216–228, 279–283.

53. L. Schlesinger, Über Gauss' Jugendarbeiten zum arithmetisch-geometrischen Mittel, Jber. Deutsch. Math.-Verein., 20 (1911) 396–403.

54. D. Shanks, Review 15, Math. Comp., 31 (1977) 1044.

55. ______, Quartic approximations for π, Abstract No. 80T-A183, Abstracts Amer. Math. Soc., 1 (1980) 558.

56. D. Shanks and J. W. Wrench, Jr., Calculation of π to 100,000 decimals, Math. Comp., 16 (1962) 76–99.

57. C. L. Siegel, Topics in Complex Function Theory, vol. 1, Wiley-Interscience, New York, 1969.

58. H. C. Thacher, Jr., Iterated square root expansions for the inverse cosine and inverse hyperbolic cosine, Math. Comp., 15 (1961) 399–403.

59. J. Todd, A problem on arc tangent relations, this MONTHLY, 56 (1949) 517–528.

60. ______, The lemniscate constants, Commun. ACM, 18 (1975) 14–19, 462.

61. ______, Basic Numerical Mathematics, Academic Press, New York, 1981.

62. H. W. Turnbull, Editor, James Gregory Tercentenary Memorial Volume, published for the Royal Society of Edinburgh, London, 1939.

63. G. N. Watson, The marquis and the land-agent; a tale of the eighteenth century, Math. Gaz., 17 (1933) 5–17.

64. D. T. Whiteside, Patterns of mathematical thought in the latter seventeenth century, Archive for History of Exact Sciences, vol. 1, 1960–1962.

65. J. W. Wrench, Jr., The evolution of extended decimal approximations to π, Math. Teacher, 53 (1960) 644–650.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

DON'T TRY TO SOLVE THESE PROBLEMS!

RICHARD K. GUY

Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4

Such an exhortation will likely produce the opposite effect, but I'm serious, and I'll explain why. This article has been in mind for some time, but its eruption is triggered by a proposal from

Schmuel Schreiber, Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel.

Problem 0. For an integer a define the set S_a inductively by

$$(1)\ a \in S_a,\ (2)\ \text{if } k \in S_a, \text{ then } 2k+2 \in S_a,\ (3)\ \text{if } k \in S_a, \text{ then } 3k+3 \in S_a.$$

Equivalently, define a function $s_a(n)$ on the integers by

$$(1) \qquad s_a(1) = a, \quad (2)\ s_a(2k) = 2s_a(k) + 2, \quad (3)\ s_a(2k+1) = 3s_a(k) + 3.$$

For $a < -3$ or $a > 2$ is s_a injective? Or does S_a contain repeated elements?

Some of you are already scribbling, in spite of the warning! More cautious readers may have been reminded of other problems, perhaps one or more of the following.

Problem 1. The diophantine equation $a^2 + b^2 + c^2 = 3abc$ has the **singular solutions** $(1,1,1)$ and $(2,1,1)$. Other solutions can be generated from these, because the equation is quadratic in each variable, for example, $b = 2$, $c = 1$ gives $a^2 - 6a + 5 = 0$, $a = 1$ or 5 and $(5,2,1)$ is a solution. Each solution, apart from the singular ones, is a **neighbor** of just three others, and they form a binary tree. Is this a genuine tree, or can the same number be generated by two different routes through it?

Problem 2. Consider the sequence $a_{n+1} = a_n/2$ (a_n even), $a_{n+1} = 3a_n + 1$ (a_n odd). For each positive integer a_1 is there a value of n such that $a_n = 1$?

Problem 3. Consider the mapping

$$2m \to 3m, \quad 4m - 1 \to 3m - 1, \quad 4m + 1 \to 3m + 1.$$

This generates the cycles (1), (2,3), (4,6,9,7,5) and (44,66,99,74,111,83,62,93,70,105,79,59). Are there others?

Problem 0 can be visualized as a binary tree generated by the pair of **unary** functions $a \to 2a + 2$, $a \to 3a + 3$. For example, if $a = 1$, we have Figure 1.

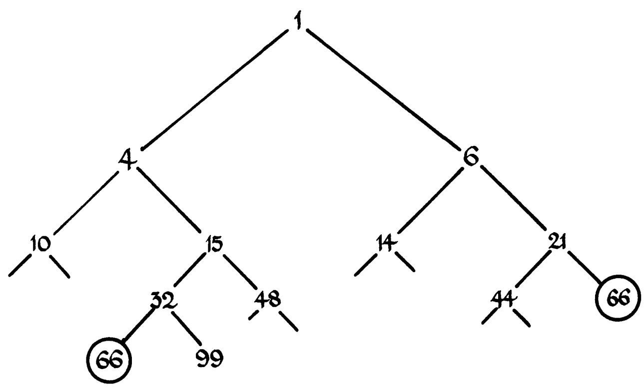


FIG. 1. Binary tree generated by two unary functions.

The number 66 appears twice in Figure 1, by making three steps to the right, or by making one to the left, one to the right and two to the left. A right step multiplies by 3 (roughly); a left step multiplies by 2 (roughly); the coincidence is roughly explained by the approximation: two right $\approx$ three left; $3^2 \approx 2^3$. Is this another example of the strong law of small numbers [**11**]? If we try other small values for a, we find similar coincidences (Figure 2).

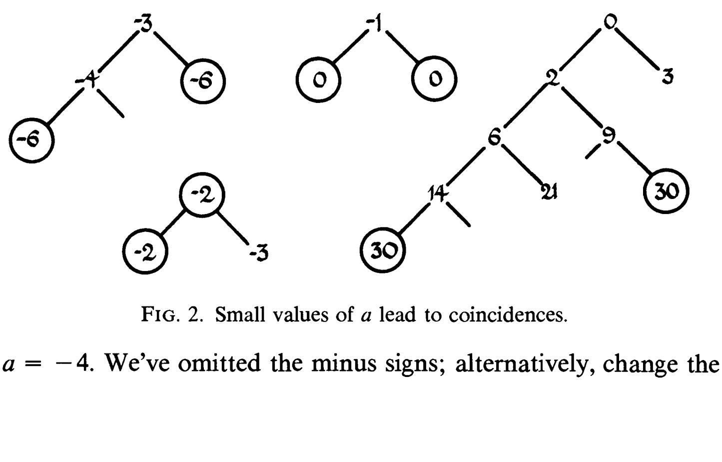


FIG. 2. Small values of a lead to coincidences.

Let us look at $a = -4$. We've omitted the minus signs; alternatively, change the plus signs to

minuses in each of conditions (2) and conditions (3) in Problem 0. The binary tree is now as in Figure 3.

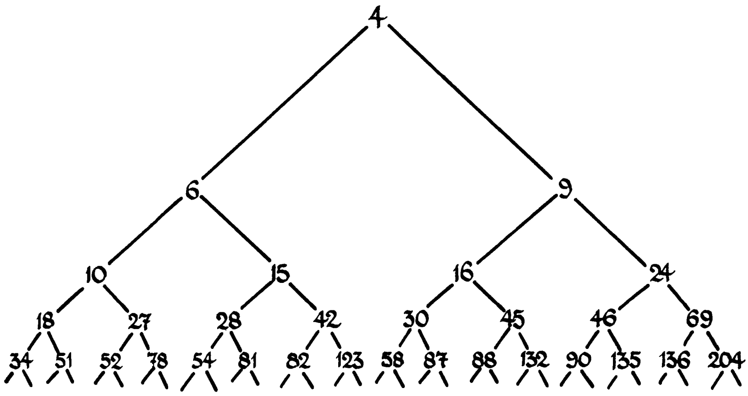


FIG. 3. Binary tree generated by $a \to 2a - 2$ and $a \to 3a - 3$.

The numbers that appear, when arranged in numerical order, are

4, 6, 9, 10, 15, 16, 18, 24, 27, 28, 30, 34, 42, 45, 46, 51, 52, 54, 58, 66, 69, 78, 81, 82, 87, 88, 90, 99, 100, 102, 106, 114, 123, 130, 132, 135, 136, 150, 153, 154, 159, 160, 162, 171, 172, 174, 178, 195, 196, 198, 202, 204, 210,....

Does a number ever occur twice? I sent an earlier draft of this article to John Leech, who disobeyed orders, started scribbling, and found that had I continued Figure 3, I would have discovered 258 and 402 on the next row, with both of these repeated two rows further down. In fact, reinstating the minus signs and using L for a left step and R for a right one, we can write these coincidences as

$$-4L^7 = -258 = -4RL^2R^2 \quad \text{and} \quad -4L^2RL^4 = -402 = -4R^2LR^2$$

and those starting from 1 and 0 in Figures 1 and 2 as

$$1LRL^2 = 66 = 1R^3 \quad \text{and} \quad 0L^4 = 30 = 0LR^2.$$

The last coincidence starts with a left step in either case, so can be regarded as a coincidence starting with 2: $2L^3 = 30 = 2R^2$. Leech used his computer to find coincidences starting with 3, 4 and 5:

$$3LRLRL^{10} = 177150 = 3R^7LR^2$$
$$4L^4R^8 = 626574 = 4RL^7RLRL^4$$
$$5L^2RL^2R^6 = 241662 = 5R^2L^{12}.$$

Since $3R = 12 = 5L$ and $3L^2 = 18 = 5R$, this last gives another coincidence starting from 3. We have seen coincidences starting from -1, -2, -3 and -4; Leech also found

$$-5L^3R^4 = -1986 = -5R^2L^6$$
$$-7LRL^2RL^3RL^4 = -143250 = -7R^2L^2R^6$$
$$-8LRL^8 = -9474 = -8R^4LR^2$$
$$-9L^{10} = -7170 = -9RL^2R^4$$
$$-10L^3RL^6 = -12354 = -10R^2LR^4$$
$$-11L^5R^2L^6 = -166146 = -11R^4LRL^2R^2$$

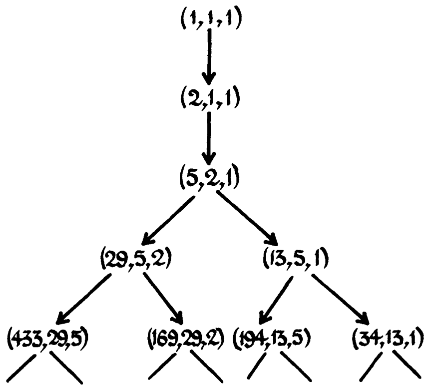


FIG. 4. Binary tree of Markoff triples.

and conjectures that there will always be "duplicate fruits on the trees" no matter where you start. If you want a coincidence starting with -6, note that $-6R = -10$, but it would be cleaner to find one of the form $-6L\ldots = -6R\ldots$ and I'm tempted to strengthen Leech's conjecture to:

¿ For each integer n, there are strings $L\ldots$ and $R\ldots$ such that $nL\ldots = nR\ldots$?

Stop press! In yet another letter, Leech proves his conjecture, so don't bother to solve this problem!

In problem 1 a binary tree is similarly generated by the pair of ternary functions

$$(a, b, c) \to (3ab - c, a, b), \quad (a, b, c) \to (3ac - b, a, c)$$

as in Figure 4.

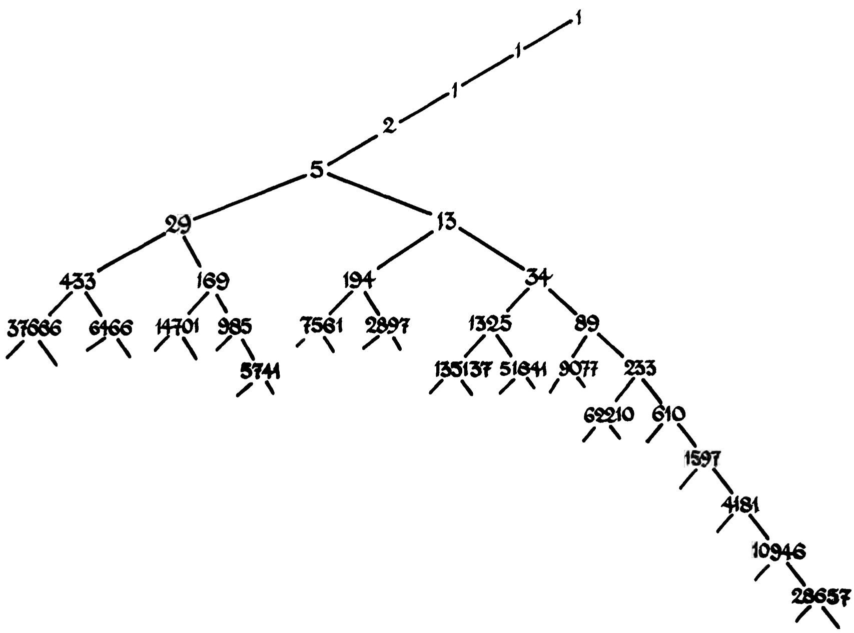


FIG. 5. Simplified Markoff tree.

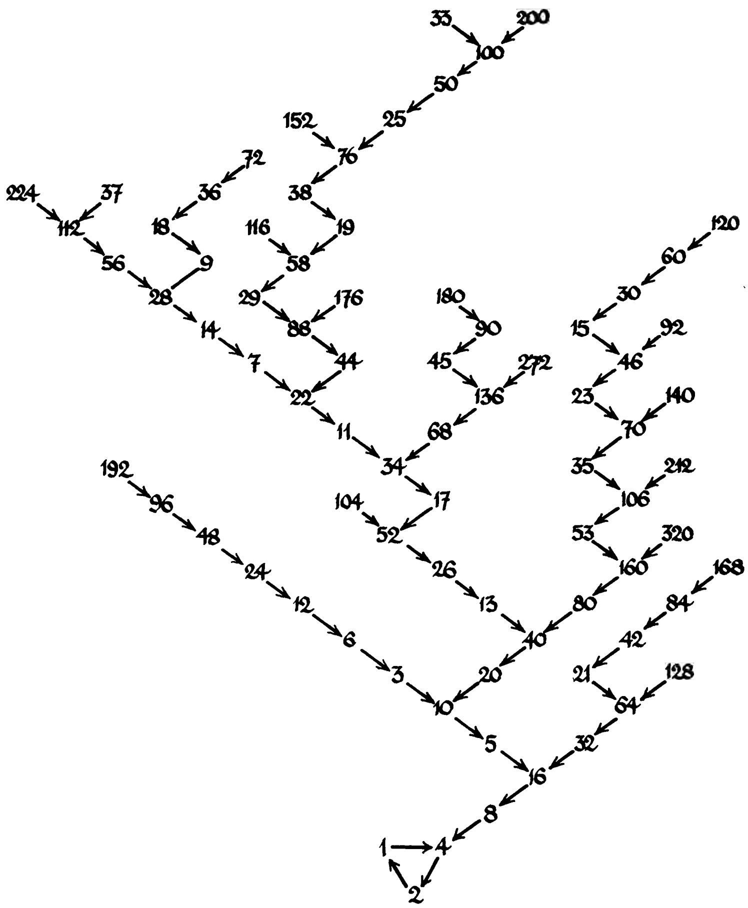


FIG. 6. Does the Collatz algorithm give any more cycles?

We can exhibit more of the tree by simplifying it as in Figure 5. To recapture the triples from this, choose any entry for a, and its immediate predecessor for b. Then c is found when travelling up the tree, just after the first step after the first rightward step. For example, $a = 985$ has predecessor $b = 169$. When travelling upwards from 985, the first rightward step is from 29 to 5. The next step is from 5 to 2, so $c = 2$.

Whether or not there are repetitions in the sequence of **Markoff numbers**

$$1,2,5,13,29,34,89,169,194,233,433,610,985,1325,1597,2897,$$
$$4181,5741,6466,7561,9077,10946,14701,28657,37666,51641,\ldots$$

has become a notorious problem. There are occasional claims to have proved uniqueness, but none seem to hold water [**19**]. Don Zagier [**23**] has some results on distribution, but none on distinctness. He can show that the problem is equivalent to the unsolvability of a certain system of diophantine equations, so we may be nearing the realm of Hilbert's tenth problem. Hence the title of this paper.

Problem 2 is associated with various names: Collatz, Hasse, Kakutani, Syracuse. It is just as notorious. Lothar Collatz told me that he thought of it when a student. One of its several waves of popularity started when he mentioned it to several people at the 1950 International Mathematical Congress in Cambridge (the wrong Cambridge). Presumably some mathematicians from Syracuse (the wrong Syracuse) became interested in it; the boys from Syracuse can perhaps fill in that bit of history.

Is the graph of the Collatz sequence unicyclic? Figure 6 includes all the numbers up to 26; the branch containing 27 is a much longer one, but still comes down to 1 after 111 steps.

After a long and inconclusive correspondence some years ago, a claimant to have a proof eventually admitted that "Erdős says that mathematics is not yet ripe enough for such questions." Hence the title of this paper.

Problem 3 is one of John Conway's **permutation sequences**. It is similar to the Collatz problem, but here the function has an inverse

$$3m \to 2m, \quad 3m - 1 \to 4m - 1, \quad 3m + 1 \to 4m + 1$$

(if the number's a multiple of 3, take a third off; otherwise add a third on) so the sequence can be pursued in either direction. Its graph consists of a number of disjoint cycles and doubly infinite chains. But it hasn't even been proved that an infinite chain exists! What is the status of the sequence containing the number 8?

$$\ldots, 41, 31, 23, 17, 13, 10, 15, 11, 8, 12, 18, 27, 20, 30, 45, 34, 51, 38, 57, 43, 32, \ldots.$$

What gives a cycle? Each term is either $3/2$ times the previous one, or approximately $3/4$ of it. Our best chance of getting back to an earlier value is to find a power of 3 which is close to a power of 2. The known cycles of lengths 1, 2, 5 and 12 correspond to the approximations of 3^1, 3^2, 3^5 and 3^{12} by 2^2, 2^3, 2^8 and 2^{19}. The last is the ratio of D sharp to E flat! In fact in each of problems 0, 2 and 3, the convergents

$$\frac{1}{1} \quad \frac{2}{1} \quad \frac{3}{2} \quad \frac{8}{5} \quad \frac{19}{12} \quad \frac{65}{41} \quad \frac{84}{53} \quad \frac{485}{306} \quad \frac{1054}{665} \quad \frac{24727}{15601} \quad \frac{50508}{31867} \quad \cdots$$

to the continued fraction for $\log 3$ to the base 2 are of significance. Note that there are cycles corresponding to the denominators 1, 2, 5 and 12. It has been shown that there are none of length 41, 53 or 306. Computers can push numerical results quite a long way, but it's not clear that they can be of any use with such problems.

In [5] Conway relates families of sequences similar to that in Problem 3 to the vector reachability problem and Minsky machines. Hence the title of this paper.

Postscript. After I'd written this, the following problem of David A. Klarner reached me from two different directions. There seem to be some significant similarities.

Problem 4. Let S be the smallest set of numbers such that $1 \in S$ and if $x \in S$, then $2x$, $3x + 2$ and $6x + 3$ each belong to S, i.e.,

$$S = \{1, 2, 4, 5, 8, 9, 10, 14, 15, 16, 17, 18, 20, 26, 27, 28, 29, 30, 32, 33, 34, \ldots\}.$$

Does S have positive density? That is, as $n \to \infty$, is

$$\liminf |S \cap \{1, 2, \ldots, n\}|/n > 0?$$

References

1. Michael Beeler, William Gosper, and Rich Schoeppel, Hakmem, Memo 239, Artificial Intelligence Laboratory, M.I.T., 1972, p. 64.

2. Corrado Böhm and Giovanna Sontacchi, On the existence of cycles of given length in integer sequences like $x_{n+1} = x_n/2$ if x_n even, and $x_{n+1} = 3x_n + 1$ otherwise, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fiz. Mat. Natur. (8) 64 (1978) 260–264.

3. J. W. S. Cassels, An Introduction to Diophantine Approximation, Cambridge, 1957, 27–44.

4. H. Cohn, Approach to Markoff's minimal forms through modular functions, Ann. of Math., Princeton (2) 61(1955) 1–12.

5. J. H. Conway, Unpredictable iterations, Proc. Number Theory Conf., Boulder, 1972, 49–52.

6. R. E. Crandall, On the "$3x + 1$" problem, Math. Comput., 32 (1978) 1281–1292; MR **58** #494; Zbl. 395.10013.

7. L. E. Dickson, Studies in the Theory of Numbers, Chicago Univ. Press, 1930, Chap. 7.

8. C. J. Everett, Iteration of the number-theoretic function $f(2n) = n, f(2n + 1) = 3n + 2$, Adv. in Math., 25 (1977) 42–45; MR **56** #15552; Zbl. 352.10001.

9. G. Frobenius, Über die Markoffschen Zahlen, S.-B. Preuss. Akad. Wiss. Berlin (1913) 458–487.

10. Martin Gardner, Mathematical Games, A miscellany of transcendental problems, simple to state but not at all easy to solve, Scientific Amer., 226 # 6(Jun 1972) 114–118, esp. p. 115.

11. Martin Gardner, Mathematical Games, Patterns in primes are a clue to the strong law of small numbers, Scientific Amer., 243 #6 (Dec 1980) 18–28.

12. Richard K. Guy, Unsolved Problems in Number Theory, Springer, New York, 1981, Problems D12, E16, E17.

13. E. Heppner, Eine Bemerkung zum Hasse-Syracuse-Algorithmus, Arch. Math. (Basel), 31 (1977/79) 317–320; MR **80**d: 10007; Zbl. 377.10027.

14. David C. Kay, Pi Mu Epsilon J., 5 (1972) 338.

15. A Markoff, Sur les formes quadratiques binaires indéfinies, Math. Ann., 15 (1879) 381–409.

16. H. Möller, Über Hasses Verallgemeinerung der Syracuse-Algorithmus (Kakutani's Problem), Acta. Arith., 34 (1978) 219–226; MR **57** #16246; Zbl. 329.10008.

17. R. Remak, Über indefinite binäre quadratische Minimalformen, Math. Ann., 92 (1924) 155–182.

18. R. Remak, Über die geometrische Darstellung der indefinitiven binären quadratischen Minimalformen, Jber. Deutsch Math.-Verein, 33 (1925) 228–245.

19. Gerhard Rosenberger, The uniqueness of the Markoff numbers, Math. Comp., 30 (1976) 361–365; but see MR **53** #280.

20. Ray P. Steiner, A theorem on the Syracuse problem, Congressus Numerantium XX, Proc. 7th Conf. Numerical Math. Comput. Manitoba, 1977, 553–559; MR **80**g:10003.

21. Riho Terras, A stopping time problem on the positive integers, Acta Arith., 30 (1976) 241–252; MR 58 #27879 (and see 35 (1979) 100–102; MR **80**h:10066).

22. L. Ja. Vulah, The diophantine equation $p^2 + 2q^2 + 3r^2 = 6pqr$ and the Markoff spectrum (Russian), Trudy Moskov. Inst. Radiotehn. Elektron. i Avtomat. Vyp. 67 Mat. (1973) 105–112, 152; MR **58** #21957.

23. Don B. Zagier, Distribution of Markov numbers, Abstract 796–A37, Notices Amer. Math. Soc., 26 (1979) A-543.

24. David A. Klarner, An algorithm to determine when certain sets have 0-density, J. Algorithms, 2 (1981) 31–43; Zbl. 464.10046.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, NM 55057.

WELL-DISTRIBUTED MEASURABLE SETS

WALTER RUDIN
Department of Mathematics, University of Wisconsin, Madison, WI 53706

THEOREM. *There is a measurable set $A \subset I = [0, 1]$ such that*

$$0 < m(A \cap V) < m(V)$$

for every nonempty open set $V \subset I$.

Proof. Let CTDP mean: Compact Totally Disconnected subset of I, having Positive measure.

Let $\{I_n\}$ be an enumeration of all segments in I whose endpoints are rational. Construct sequences $\{A_n\}$, $\{B_n\}$ of CTDP's as follows:

Start with disjoint CTDP's A_1 and B_1 in I_1.

Once $A_1, B_1, \ldots, A_{n-1}, B_{n-1}$ are chosen, their union C_n is CTD, hence $I_n \setminus C_n$ contains a

C E N T E R   S E C T I O N
(Vol. 90, No. 1, January 1983)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S(13-14), L. Two-Year College Mathematics Readings. Ed: Warren Page. MAA, 1981, vii + 304 pp, $19.50 (P). [ISBN: 0-88385-435-X] 45 short notes arranged in eight sections--algebra, geometry, number theory, calculus, probability and statistics, calculators and computers, mathematics education, recreational mathematics. Contains a wealth of brief mathematical topics suitable for class projects and lecture enrichment. A useful resource for anyone teaching elementary mathematics, whether at a two-year college or a university. LAS

General, S(15-16), L. La mystification mathématique. Alain Bouvier. Hermann, 1981, 158 pp, 62F (P). [ISBN: 2-7056-1403-6] An interesting excursion into mathematics, discussion of the nature of mathematical activity, mathematical proof, the influence of testing on teaching, and mathematical pedagogy. Reminiscent of The Mathematical Experience (and clearly influenced by the writings of P. Davies) but lacking the scope and depth of that work. SG

General, P. The File: Case Study in Correction (1977-1979). Serge Lang. Springer-Verlag, 1981, xi + 712 pp, $19.80 (P). [ISBN: 0-387-90607-X) Reproduction of documents (letters, reports, published articles) concerning Lang's dispute with E.C. Ladd and S.M. Lipset about the scholarly and statistical standards employed in their widely reported Survey of the American Professoriate. Includes all relevant source documents, photocopied in chronological order, together with introduction and rebuttals by the protagonists. In addition to its value as a record of an important academic dispute, it contains much of interest to anyone dealing with social science research methodology. LAS

General, P. Lecture Notes in Mathematics-915: Categorical Aspects of Topology and Analysis. Ed: B. Banaschewski. Springer-Verlag, 1982, xi + 385 pp, $20 (P). [ISBN: 0-387-11211-1] Proceedings of an August 1981 conference held at Carleton University, Ottawa. LAS

Elementary, S. How to Think with Numbers. Robert L. Hershey. William Kaufmann, 1982, 133 pp, $7.95 (P). [ISBN: 0-86576-014-4] A breezy, simplistic exhortation to think carefully about percentages, interest, probabilities, and monetary decisions. Many easy examples; reference tables with compound interest, annuities and future value. Perhaps useful as a remedy for math-avoidance. LAS

Mathematics Appreciation, T(13: 1). Mathematics. Howard L. Rolf. Allyn & Bacon, 1982, x + 486 pp, $22.95. [ISBN: 0-205-07627-0] Chapters on mathematical models, algorithms, and applications (to interest, depreciation, insulation, etc.) precede the presentation of elementary concepts from algebra, geometry, set theory, probability and statistics, graph theory and number theory. Optional sections introduce flow charts and Basic programs. JNC

Mathematics Appreciation, T*(11-16: 1, 2), L. Mathematics: A Human Endeavor, Second Edition. Harold R. Jacobs. WH Freeman, 1982, xiii + 649 pp, $17.95. [ISBN: 0-7167-1326-8] A careful, thorough revision of this classic text (TR, First Edition, August-September 1971; ER, August-September 1972). Major change: sections on "Problems for Further Exploration" conclude each chapter, giving several extended activity or discussion projects. LAS

Mathematics Appreciation, S(14). Magic Cubes: New Recreations. William H. Benson, Oswald Jacoby. Dover, 1981, 142 pp, $4 (P). [ISBN: 0-486-24140-8] Cyclical method for magic squares is extended to three dimensions. Some previously unconstructed cubes are presented, e.g., the first double-even-order perfect magic cube of order 12. A complete survey with many examples and detailed instructions. JRG

Precalculus, S(13). Essential Mathematics for College Physics, A Self Study Guide. Michael Ram. Wiley, 1982, 278 pp, $9.95 (P). [ISBN: 0-471-86454-4] This photo-offset edition is designed as a self-study guide to be used in conjunction with a standard text for a non-calculus-based physics course. Topics include scientific notation, algebra, geometry, analytical geometry, trigonometry, vectors, exponential and logarithmic functions and arithmetic and geometric sequences. JNC

Precalculus, T(13: 1), S, L. College Algebra, Second Edition. Karl J. Smith, Patrick J. Boyle. Brooks/Cole Pub, 1982, xi + 444 pp, $19.95. [ISBN: 0-8185-0489-7] Covers the usual precalculus material (emphasizing functions) and more: matrices and determinants, introduction to infinite series, combinatorics and probability. Focus on the techniques and logic of problem-solving in sections on word problems and in a full chapter on mathematical proof. Exercises are graded at three levels of difficulty. PZ

Education, P, L. Films in the Mathematics Classroom. Barbara J. Bestgen, Robert E. Reys. NCTM, 1982, vi + 90 pp, $6.75 (P). [ISBN: 0-87353-195-7] Reviews of nearly 250 mathematics films for K-12, together with introductory essays on effective use of films. Includes lists of distributors and lists of films in series, only some of which are reviewed in the pamphlet. A useful complement to David Schneider's Annotative Bibliography of Films and Videotapes for College Mathematics (TR, November 1981). LAS

Education, T. Mathematics: A Good Beginning: Strategies for Teaching Children, Second Edition. Andria P. Troutman, Betty K. Lichtenberg. Brooks/Cole Pub, 1982, x + 502 pp, $17.95 (P). [ISBN: 0-8185-0492-7] For pre-service and in-service teachers of grades K-6; describes mathematical ideas which children need to learn; ways to present these ideas and difficulties children may encounter while learning them. (TR, First Edition, January 1978.) JNC

History, S(14-18), P, L.** A History of Greek Mathematics. Sir Thomas Heath. Dover Pub, 1981, $8.50 (P) each. [ISBN: 0-486-24073-8] Volume I: From Thales to Euclid, xv + 446 pp; Volume II: From Aristarchus to Diophantus, xi + 586 pp. [ISBN: 0-486-24074-6] Republication of Heath's masterful work (1921) at a very reasonable price. Order is roughly chronological, but with special chapters devoted to the history of certain important problems (e.g., doubling of the cube) and to the roles of major figures (Euclid, Archimedes). Noted for careful scholarship. JRG

History, T(15-17), S*, P, L.** From the Calculus to Set Theory, 1630-1910: An Introductory History. Ed: I. Grattan-Guinness. Gerald Duckworth & Co, 1980, 306 pp. [ISBN: 0-7156-1295-6] An introductory history, arranged in six separately-authored chapters written by historians of different mathematical periods. The intent is more pedagogical than scholarly, to introduce undergraduate mathematics majors to the important influences that transformed Newtonian calculus into modern analysis. An excellent seminar text, or background resource for anyone teaching analysis. LAS

History, L*. Collected Papers on Wave Mechanics. E. Schrödinger. Chelsea Pub, 1982, xiii + 207 pp, $14.95. [ISBN: 0-8284-1302-9] Third English edition of the 1927 original German volume, augmented with the text of "Four Lectures on Wave Mechanics," originally published in Glasgow in 1928. LAS

History, T(15-16: 1), S*, L.** Great Moments in Mathematics (After 1650). Howard Eves. Dolciani Math. Expos., No. 7. MAA, 1981, xii + 263 pp, $23.50. [ISBN: 0-88385-307-8] Sequel to Great Moments in Mathematics (Before 1650) (TR, March 1982) providing twenty lectures from the modern (calculus) era. Topics include the discovery of Fourier series, non-Euclidean geometry, non-commutative algebra, transfinite numbers, and abstract spaces. A very readable approach to history, touching on interesting highlights without overwhelming detail. Exercises, with hints and answers at the back and a thorough index make the book well suited to class use. LAS

History, P, L*. The Scottish Book: Mathematics from the Scottish Café. Ed: R. Daniel Mauldin. Birkhauser Boston, 1981, xiii + 268 pp, $24.95. [ISBN: 3-7643-3045-7] From 1935 until occupation in 1941, the mathematics circle of Lwow, Poland (including Banach, Ulam, Steinhaus, Mazur, Kac) recorded in a notebook at their favorite café important, difficult problems in set theory, topology, measure theory and functional analysis. This unique notebook is reproduced here (in English translation) with occasional commentaries and modern references. The volume is introduced by brief anecdotal essays about the origins of the Scottish book by Ulam, Kac, Zygmund and Erdös. LAS

Foundations, P, L. Friedrich Waismann, Lectures on the Philosophy of Mathematics. Ed: Wolfgang Grassl. Humanities Pr, 1982, 170 pp, $17 (P). [ISBN: 90-6203-983-9] Six essays from the 50's by the man who in the 30's was chiefly responsible for bringing Wittgenstein's philosophy to the attention of the Vienna Circle. Editor's introduction traces the collaboration and falling out of the two men and outlines the central ideas of their shared constructivist conception of mathematics, which challenges all three of the then dominant schools. GHM

Number Theory, T(16-17), S, P*, L*. Introduction to Number Theory. Loo-Keng Hua. Trans: Peter Shiu. Springer-Verlag, 1982, xviii + 572 pp, $46. [ISBN: 0-387-10818-1] A translation and update of the author's 1956 text. A monumental work which is reminiscent of Hardy and Wright in both style and content. Contains occasional exercises. A very impressive book. CEC

Number Theory, P. Lecture Notes in Mathematics-938: Number Theory. Ed: K. Alladi. Springer-Verlag, 1982, ix + 177 pp, $10.70 (P). [ISBN: 0-387-11568-4] A representative collection of papers presented at the Third Matscience Conference on Number Theory held at Mysore, India during June 3-6, 1981. Includes a talk by Erdös and problems presented at a problems session. CEC

Number Theory, T*(16-17: 1), S, P, L*. A Classical Introduction to Modern Number Theory. Kenneth Ireland, Michael Rosen. Grad. Texts in Math., No. 84. Springer-Verlag, 1982, xiii + 341 pp, $28. [ISBN: 0-387-90625-8] A revised and expanded version of the author's Elements of Number Theory which was published in 1972 (TR, August-September 1972). A familiarity with undergraduate level abstract algebra is assumed. Unlike most introductory number theory texts, the focus is on topics

which point to algebraic number theory and arithmetic algebraic geometry. CEC

Linear Algebra, T(14: 1). Elementary Linear Algebra, Third Edition. Bernard Kolman. Macmillan Pub, 1982, xii + 356 pp, $23.95. [ISBN: 0-02-365990-4] New features of the Third Edition include: more examples and exercises (at all levels); treatment of independence and basis in separate section. This edition bears a strong resemblance to the non-applications portions of Introductory Linear Algebra with Applications, Second Edition, by the same author. Note, however, in the text being reviewed, the use of Greek letters, and the integration of theoretical exercises into the rest of the exercise section. (TR, First Edition, June-July 1970 and January 1971, Extended Review, March 1974; TR, Second Edition, April 1977.) JRG

Linear Algebra, T(14: 1). Linear Algebra, A Concrete Introduction. Dennis M. Schneider, Manfred Steeg, Frank H. Young. Macmillan Pub, 1982, xi + 347 pp, $21.95. [ISBN: 0-02-476810-3] An elementary, applied approach, beginning with systems of equations and Gaussian elimination, moving through inconsistent systems (least squares, projections) to linear transformations, eigenvalues, and numerical methods. Both the approach and organization are like Strang's Linear Algebra and Its Applications, but the exercises and discussion are more suited to average students. Excellent collection of applications, well integrated into the text. LAS

Algebra, T(16-18: 1),S, P, L. Boolean Matrix Theory and Applications. Ki Hang Kim. Pure & Appl. Math., V. 70. Dekker, 1982, xiv + 288 pp, $38.50. [ISBN: 0-8247-1788-0] A fascinating introduction to the theory and some of the applications of Boolean matrices (matrices with entries 0,1 using Boolean operations). The discussion proceeds from basic algebra through the notions of rank, generalized inverse, combinatorial properties, and asymptotic forms with numerous excursions into applications. Many exercises, references, index. JS

Algebra, P. Lecture Notes in Mathematics-907: Dualisierende Komplexe in der lokalen Algebra und Buchsbaum--Ringe. Peter Schenzel. Springer-Verlag, 1982, vii + 161 pp, $9.60 (P). [ISBN: 0-387-11187-5]

Algebra, P. Rings, Modules, and Preradicals. L. Bican, T. Kepka, P. Němec. Pure & Appl. Math., V. 75. Dekker, 1982, ix + 241 pp, $35 (P). [ISBN: 0-8247-1568-3] An introduction to the research of the authors (with P. Jambor) conducted over the past ten years. The general theme is the relationship between the structure of a ring R and properties of preradicals for the category of left R-modules. SG

Algebra, S*(18), P*, L*. Finite Simple Groups: An Introduction to Their Classification. Daniel Gorenstein. Plenum Pr, 1982, x + 333 pp, $29.50. [ISBN: 0-306-40779-5] An expanded version of the author's 1979 Bulletin article. The book opens with an overview of the recently completed classification of finite simple groups. Two chapters are then devoted to a description of the known simple groups; the final chapter presents the local techniques used in the classification proof. A beautiful (and we hope precedent-setting) model of mathematical exposition. SG

Algebra, T(17), S, P, L. Between Nilpotent and Solvable. Ed: Henry G. Bray, et al. Polygonal Pub, 1982, vi + 231 pp, $22. [ISBN: 0-936428-06-6] An exposition of the theory of finite solvable groups. Chapters include supersolvable groups; M groups; CLT and non-CLT groups; miscellaneous classes. Intended for a general audience of mathematicians who have already had a course in group theory. JG

Algebra, P. Commutative Algebra: Analytical Methods. Ed: Richard N. Draper. Pure & Appl. Math., V. 68. Dekker, 1982, viii + 291 pp, $38.50 (P). [ISBN: 0-8247-1282-X] Contributed papers from an August 1979 CBMS Regional Research Conference at George Mason University. (The central lectures of this conference, by Melvin Hochster, will be published separately by the American Mathematical Society.) LAS

Algebra, T(18), S, P. The Representation Theory of Finite Groups. Walter Feit. Math. Lib., V. 25. Elsevier North-Holland, 1982, xiv + 502 pp, $55. [ISBN: 0-444-86155-6] A comprehensive treatment of the theory of modular and p-adic representations of finite groups as it exists at present. It contains most of what is known about representations of finite groups in general. Much of the material has never before appeared in print. LCL

Algebra, T*(15-17: 1), S, L*. A First Course in Abstract Algebra, Third Edition. John B. Fraleigh. Addison-Wesley, 1982, xviii + 478 pp, $23.95. [ISBN: 0-201-10405-9] The sections on algebraic topology have been eliminated in this edition (TR, First Edition, August-September 1969). Material on group actions, Burnside counting and the Sylow theorems has been added. There are additional exercises, and answers to even numbered exercises have been omitted. A solid introduction to groups, rings and fields. CEC

Algebra, P. Lecture Notes in Mathematics-931: Finite Rank Torsion Free Abelian Groups and Rings. David M. Arnold. Springer-Verlag, 1982, vii + 191 pp, $10.70 (P). [ISBN: 0-387-11557-9] An exposition of results obtained in the last decade in the theory of finite rank torsion free abelian groups by Hunter, Lady, Richman, Walker, Warfield, the author and others. SG

Algebra, T(17-18: 1), S, P, L. Group Theory I. Michio Suzuki. Grund. der math. Wissenschaften, B. 247. Springer-Verlag, 1982, xiv + 434 pp, $48. [ISBN: 0-387-10915-3] First of two volumes in the English translation from the original Japanese version; completed before much of the recent dramatic progress in work on finite simple groups, but still quite useful. Chapter 1 is a quick look at

basic concepts; Chapter 2 is largely on major theories including those of Sylow, Krull-Remak-Schmidt, Schur-Zassenhaus. Chapter 3 deals with special classes: Abelian, symmetric, Coxeter, finite simple. Exercises, bibliography, index. JS

Algebra, S(18), P. Universal Algebra. Ed: B. Csákány, E. Fried, E.T. Schmidt. Elsevier North-Holland, 1982, 804 pp, $151. [ISBN: 0-444-85405-3] A collection of 68 papers, all variants of talks presented at a colloquium on universal algebra in June 1977. Organized into three general areas: 1) pure universal algebra; 2) applications in computer science; and 3) lattice theory. Inordinately priced. JS

Algebra, S(17-18), P. Elements of Group Theory for Physicists, Third Edition. A.W. Joshi. Halsted Pr, 1982, xiii + 334 pp, $16.95. Changes from earlier editions include a flow-chart for an algorithmic approach to determining character tables together with examples. An appendix on mapping has been added. (TR, First Edition, March 1975.) JS

Algebra, T(18: 1), S, P. Lectures on Representations of Complex Semi-Simple Lie Groups. Thomas J. Enright. Springer-Verlag, 1981, 91 pp, $7.10 (P). [ISBN: 0-387-10829-7] Based on a lecture series, the aim is to describe a functorial correspondence relating admissible representations for a complex semi-simple Lie group and highest weight modules for the Lie algebra. Based largely on work of Zelobenko and Langlands. Bibliography. JS

Calculus, T(13-14: 2, 3). Mathematical Analysis: Business and Economic Applications, Fourth Edition. Jean E. Weber. Harper & Row, 1982, xii + 719 pp. [ISBN: 0-06-046977-3] A comprehensive introduction to calculus (through Kuhn-Tucker conditions and double integrals), differential and difference equations, matrix algebra (through characteristic vectors), linear programming, game theory, and Markov processes. Each technique is supported by numerous routine exercises and by extensive applications to business and economic theory. (TR, Third Edition, December 1976.) LAS

Calculus, T*(15-16: 1, 2), S, L. Advanced Calculus with Applications. Nicholas J. De Lillo. Macmillan Pub, 1982, xii + 836 pp, $31.95. [ISBN: 0-02-328220-7] Informal and readable treatment of the calculus of one and several variables, bridging the gap between the usual calculus sequence and courses in real and complex analysis. Modern treatment of differential calculus of transformations. Required linear algebra is included. Sufficient material for two semesters. Good exercises complement and extend text. Applications, none surprising, both theoretical and applied, are scattered throughout. JK

Real Analysis, T(16-17: 1), S, P, L. Advanced Engineering Analysis. J.N. Reddy, M.L. Rasmussen. Wiley, 1982, xiv + 488 pp, $39.95. [ISBN: 0-471-09349-1] Written primarily for engineers and applied scientists, the book is a rather unusual (but successful) blend of classical topics (vector and tensor analysis) and abstract functional analysis (normed linear spaces, operators, Hilbert space). Application is made of both in a final chapter on calculus of variations and variational methods. Exercises, references, index. JS

Real Analysis, T(16-17: 1), S, P, L. A Second Course on Real Functions. A.C.M. Van Rooij, W.H. Schikhof. Cambridge U Pr, 1982, xiii + 200 pp, $29.50; $22.95 (P). [ISBN: 0-521-23944-3; 0-521-28361-2] A rigorous, largely self-contained look at some of the deeper classical theory with the emphasis placed on real as opposed to abstract functional analysis. Includes set theory, continuity, differentiation, measure theory, and integration (Lebesgue, Stieltjes, Perron). Many examples and counterexamples (including an index), exercises, appendices. JS

Real Analysis, P. Lecture Notes in Mathematics-937: Intégrales Exponentielles: Développements Asymptotiques, Propriétés Lagrangiennes. Edmond Combet. Springer-Verlag, 1982, viii + 114 pp, $8.80 (P). [ISBN: 0-387-11566-8] The objects of study are real n-dimensional integrals whose integrands are compactly supported functions with an exponential factor. The argument of the exponential factor has a positive real parameter--the asymptotic development of the integral as a function of this parameter is obtained. Later, a Lagrangian theory of such integrals is given, together with examples from Hamiltonian systems and variational calculus. PZ

Real Analysis, T(14-16: 1), S, L. Infinite Processes: Background to Analysis. A. Gardiner. Springer-Verlag, 1982, ix + 306 pp, $28. [ISBN: 0-387-90605-3] Unusual, lively text critically examines the why and how of infinite processes as a prologue or accompaniment to the study of analysis. Treats rational and irrational numbers, the relationship between geometry and number, and the evolution of the modern concept of a function. Many examples illustrate the subtlety and significance of the "jump" from finite to infinite. Historical perspective included. Many (interesting) exercises. JRG

Complex Analysis, S(18), P. Topics in Polynomial and Rational Interpolation and Approximation. Richard S. Varga. Pr U Montreal, 1982, 136 pp, $13 (P). [ISBN: 2-7606-0573-6] Seven self-contained chapters, each with its own references, on polynomial and rational approximation and interpolation in one complex variable. A recurrent theme is numerical computation--the author asserts that it can "sharpen...ideas and stimulate research in...complex analysis." Many open questions and conjectures. Based on 1981 NATO lectures at the University of Montreal. PZ

Complex Analysis, P*. Lectures on Approximation and Value Distribution. Tord Ganelius, Walter K. Hayman, Donald J. Newman. Pr U Montreal, 1982, 174 pp, $14 (P). [ISBN: 2-7606-0570-1] Ganelius's contribution is a treatise on rational approximation to continuous functions in the plane; Hayman

covers recent results relating value distribution to a new measure of exceptional sets; and Newman compares rational approximation methods with the fast computer approximation techniques. TAV

Differential Equations, P. Global Structural Stability of Flows on Open Surfaces. Janina Kotus, Michał Krych, Zbigniew Nitecki. Memoirs No. 261. AMS, 1982, v + 108 pp, $6. (P) [ISBN: 0-8218-2261-6]

Differential Equations, T(15-16: 1, 2). Partial Differential Equations for Scientists and Engineers. Stanley J. Farlow. Wiley, 1982, ix + 402 pp, $29.95. [ISBN: 0-471-08639-8] A practically oriented introductory text of partial differential equations organized as 47 lessons. In addition to separation of variables and integral transforms, the book discusses numerical methods, Monte Carlo methods, and several other important, but nonstandard, topics. AO

Differential Equations, P. Green's Functions and Transfer Functions Handbook. A.G. Butkovskiy. Transl: L.W. Longdon. Math. & Its Appl. Halsted Pr, 1982, 237 pp, $64.95. [ISBN: 0-470-27344-5] Differential, integral and integro-differential equations are catalogued in terms of the space dimension and orders of various derivatives, and solutions are given for each of the 24 cases considered. Technical expertise is required. A short discussion of the theory follows. TAV

Differential Equations, P. Lecture Notes in Mathematics-925: The Riemann Problem, Complete Integrability and Arithmetic Applications. Ed: D. and G. Chudnovsky. Springer-Verlag, 1982, vi + 373 pp, $20 (P). [ISBN: 0-387-11483-1] This volume contains a series of lectures on the Riemann boundary value problem which was presented at a seminar held in 1979-80 at the Institut des Hautes Etudes Scientifiques in Bures-sur-Yvette, France, and at Columbia University. CEC

Numerical Analysis, T(16-17: 1, 2), S*, P*, L. Iterative Methods for the Solution of Equations. J.F. Traub. Chelsea Pub, 1982, xviii + 310 pp, $15.95. [ISBN: 0-8284-0312-0] Textually identical with the 1964 edition which contained much new material. Rigorous treatment of the mathematical foundations of a general theory of iteration algorithms for the numerical solution of real equations and systems of real equations with emphasis on algorithm efficiency. No problems. Excellent references through 1963. Dated, but very well done. An excellent buy. JK

Numerical Analysis, T(14-16: 1). Numerical Analysis: A Practical Approach. Melvin J. Maron. Macmillan Pub, 1982, xviii + 471 pp, $26.95. [ISBN: 0-02-475670-9] The topics covered in this textbook are the standard ones (roots of equations, linear algebra, approximation and interpolation, quadrature, and ordinary differential equations). It emphasizes modern algorithms and attempts to provide an intuitive understanding of how each works. Few proofs or derivations are given. AO

Numerical Analysis, T(15-16: 1), L*. Essentials of Numerical Analysis with Pocket Calculator Demonstrations. Peter Henrici. Wiley, 1982, vi + 409 pp, $27.95. [ISBN: 0-471-05904-8] This is a very interesting new textbook which emphasizes the connections between applied mathematics and numerical analysis. A unique feature of the book is that all of the problems are designed to be carried out on a programmable pocket calculator. Most of the standard topics are covered. AO

Numerical Analysis, S(15-17), L*. Numerical Solution of Partial Differential Equations in Science and Engineering. Leon Lapidus, George F. Pinder. Wiley, 1982, 677 pp, $44.95. [ISBN: 0-471-09866-3] The numerical methods treated in this work include finite differences, finite elements, collocation, and boundary elements. After introductory chapters on the basic techniques, each type of partial differential equations is treated in a separate chapter. Each of these chapters includes an extensive bibliography. AO

Numerical Analysis, T(14-15: 1), S*, L. Numerical Methods. Robert W. Hornbeck. Quantum Pub, 1975, 310 pp, $9.95 (P). [ISBN: 0-13-626614-2] A relative bargain. Intended for science and engineering students. Includes Taylor's formula, finite differences, interpolation, approximation, nonlinear equations, linear systems, integration, ordinary differential equations, partial differential equations, matrix eigenvalue problems. Develops some of the methods. Examples, flowcharts, exercises. RWN

Functional Analysis, P. Lecture Notes in Mathematics-936: Counterexamples in Topological Vector Spaces. S.M. Khaleelulla. Springer-Verlag, 1982, xxi + 179 pp, $10.70 (P). [ISBN: 0-387-11565-X] Topological vector spaces may be Fréchet, locally convex, bornological, Montel, etc. These notes consist of approximately 200 examples of spaces or subspaces enjoying some but not others of these properties. Other such counterexamples are given in the category of topological algebras. Open problems in the same spirit are included. PZ

Functional Analysis, T(18), P. Continuous Semigroups in Banach Algebras. Allan M. Sinclair. London Math. Soc. Lect. Note Ser., No. 63. Cambridge U Pr, 1982, 145 pp, $19.95 (P). [ISBN: 0-521-28598-4] A discussion of the abstract theory of analytic one-parameter semigroups in Banach algebras. LCL

Analysis, P. Equivalence of Measure Preserving Transformations. Donald S. Ornstein, Daniel J. Rudolph, Benjamin Weiss. Memoirs No. 262. AMS, 1982, xii + 116 pp, $7.60 (P). Almost forty years ago Kakutani introduced an equivalence between ergodic measure-preserving transformations and showed its relevance in studying the measurable cross sections of measurable flows. Only three inequivalent types of equivalence of measure-preserving transformations were known until J. Feldman resurrected Kakutani's idea by constructing a fourth type. The present volume is a compilation of lectures giving a detailed exposition of Feldman's ideas and their ramifications. PH

Analysis, T*(17: 1, 2), S*, P, L. Fixed Point Theory, Volume I. James Dugundji, Andrzej Granas. PWN, 1982, 209 pp. [ISBN: 83-01-01142-4] Requires some background in topology and functional analysis. Beginning with the Banach contraction principle, it includes the major results with special attention to Leray-Schauder theory, with applications to differential and integral equations. Numerous exercises, notes and comments and a substantial bibliography. Should serve to standardize the terminology of recent results in the field. TAV

Analysis, P. Lecture Notes in Mathematics-932: Analytic Theory of Continued Fractions. Ed: W.B. Jones, W.J. Thron, H. Waadeland. Springer-Verlag, 1982, 240 pp, $12.50 (P). [ISBN: 0-387-11567-6] This volume contains the proceedings of a seminar-workshop on recent progress in the analytic theory of continued fractions held at Loen, Norway from June 5-30, 1981. CEC

Analysis, P. The Mathematics of Time: Essays on Dynamical Systems, Economic Processes, and Related Topics. Steve Smale. Springer-Verlag, 1980, vi + 151 pp, $16 (P). [ISBN: 0-387-90519-7] Reprints of 10 of Smale's papers and reviews on dynamical systems; half the volume consists of his 1967 Bulletin survey "Differentiable Dynamical Systems." Concludes with a brief, informal autobiographical note. LAS

Geometry, T(15-16), S, L**.** Convex Sets and Their Applications. Steven R. Lay. Wiley, 1982, xvi + 244 pp, $29.50. [ISBN: 0-471-09584-2] Generally set in R^n, this beautiful book is less abstract, more accessible to undergraduates than Valentine's Convex Sets with which it invites comparison. Same audience, same goals, same style as Roberts and Varberg's Convex Functions with which it is a natural companion volume. A must for every mathematics library. AWR

Geometry, T(16: 1, 2), S, L. The Mathematical Theory of Chromatic Plane Ornaments. Thomas W. Wieting. Mono. and Textbooks in Pure & Appl. Math., No. 71. Dekker, 1982, vii + 369 pp, $55. [ISBN: 0-8247-1517-9] With the goal of developing the color theory for plane ornaments, i.e., for periodic tilings of the Euclidean plane, the author begins from "scratch," developing the theory of the crystallographic groups and all of the geometry and group theory necessary to reach his goal. The only prerequisites are elementary linear algebra and group theory. SS

Algebraic Topology, P. Brown-Peterson Homology: An Introduction and Sampler. W. Stephen Wilson. CBMS Reg. Conf. Series in Math., No. 48. AMS, 1982, v + 86 pp, $9.20 (P). [ISBN: 0-8218-1699-3]

Topology, P. G Surgery II. Karl Heinz Dovermann, Ted Petrie. Memoirs No. 260. AMS, 1982, xxiii + 118 pp, $8 (P). An account of the work of the two authors which develops an equivariant surgery theory containing a generalization of the exact Wall sequence. JAS

Topology, P. Low-Dimensional Topology. Ed: R. Brown, T.L. Thickstun. London Math. Soc. Lect. Note Ser., No. 48. Cambridge U Pr, 1982, 246 pp, $27.50 (P). [ISBN: 0-521-28146-6] Papers on knots, manifolds and homotopy in 2, 3 and 4 dimensions from a 1979 conference at University College of North Wales, Bangor, U.K. With the exception of notes by Peter Scott on William Thurston's lectures on 3-manifolds, the major survey lectures of the conference are not included. LAS

Optimization, T(16-17: 3), L. Mathematical Programming and Games. Edward L. Kaplan. Wiley, 1982, xx + 588 pp, $34.95. [ISBN: 0-471-03632-3] Linear programming presented with early attention to duality and an all-integer pivoting procedure that relieves some complexity in hand calculation. Gets to the revised simplex method; last third of book deals with optimal paths in networks and the transportation problem. Accessible to a student with modest linear algebra. AWR

Optimization, T, S*, P, L.** Maxima and Minima Without Calculus. Ivan Niven. Dolciani Math. Expos., No. 6. MAA, 1981, xv + 303 pp, $24.50. [ISBN: 0-88385-306-X] A compendium of algebraic, geometric, and inequality methods for optimization problems, excluding methods of calculus, linear programming, and game theory. Treats isoperimetric problems (in two and three dimensions), tiling problems, refractions, and various problems involving inscribed geometric figures. Exercises are scattered throughout each chapter, with answers at the end. LAS

Statistics, T(14-16: 1, 2), S. Statistical Analysis: Resolving Decision Problems in Business and Management. Stephen A. Book, Marc J. Epstein. Scott Foresman, 1982, xii + 736 pp, $22.95. [ISBN: 0-673-16002-5] Descriptive techniques, probability, distributions, statistical inference, prediction and forecasting, Bayesian techniques, and nonparametric methods. Presupposes no college mathematics. Examples drawn from real situations in business. FLW

Statistics, T(13-14: 1). Introduction to Statistics, Third Edition. Ronald E. Walpole. Macmillan Pub, 1982, xv + 521 pp, $23.95. [ISBN: 0-02-424150-4] Revised, reorganized and expanded version of the author's 1974 Second Edition (TR, December 1974). Material on nonparametric statistics has been increased and placed in a separate chapter. New material also includes more descriptive statistics, sampling procedures, Bayesian methods, partial and multiple correlation, and Latin square designs. (TR, First Edition, October 1968.) RSK

Statistics, T?(14-15: 1). Elementary Statistical Methods, Third Edition. G. Barrie Wetherill. Chapman & Hall, 1982, 356 pp, $14.95 (P). [ISBN: 0-412-24000-9] Modest revision of the author's Second Edition (TR, March 1974). Main change is the addition of some simple Basic programs. It remains an old-fashioned text with a tendency toward a cookbook style. (TR, First Edition, June-July 1968.) RSK

Statistics, P. Statistics and Probability: Essays in Honor of C.R. Rao. Ed: G. Kallianpur, P.R. Krishnaiah, J.K. Ghosh. Elsevier North Holland, 1982, xi + 722 pp, $144.25. [ISBN: 0-444-86130-0] Collection of sixty-six papers covering a wide range of topics, both pure and applied. Also includes a list of all Rao's publications through 1981. Note price. RSK

Statistics, P. Prior Information in Linear Models. Helge Toutenburg. Wiley, 1982, ix + 215 pp, $39.95. [ISBN: 0-471-09974-0] In the Wiley Series in Probability and Mathematical Statistics. Uses restrictions, as well as prior estimates, to estimate coefficients in linear regression models. Concentrates on robustness of minimax estimators, restricted and two-stage least squares estimators, and mixed estimators which allow for part of the prior information to be incorrect. Aimed at both statisticians and econometricians. RSK

Statistics, P*. The Complete Categorized Guide to Statistical Selection and Ranking Procedures. Edward J. Dudewicz, Joo Ok Koo. American Sci Pr, 1982, v + 627 pp, $85 (P). [ISBN: 0-935950-03-6] Volume 6 in the American Series in Mathematical and Management Sciences. Revised and expanded version of Dudewicz' 1968 bibliography entitled A Categorized Bibliography on Multiple-Decision (Ranking and Selection) Procedures. Includes a complete collection of reviews from Mathematical Reviews and Zentralblatt für Mathematik in addition to various listings of relevant journal articles and other references. Note price. RSK

Statistics, P*. Teaching of Statistics and Statistical Consulting. Ed: Jagdish S. Rustagi, Douglas A. Wolfe. Academic Pr, 1982, xvi + 548 pp, $36. [ISBN: 0-12-604540-2] Proceedings of an international conference held at The Ohio State University in November 1980. Contains invited papers, discussions, and contributed papers in five general areas: graduate programs in statistics, teaching service courses and short courses, training statisticians for employment in industry and government, the role of statistical consulting in graduate training, and statistics at the Open University in England. RSK

Statistics, T?(13: 1). A Primer in Data Reduction: An Introductory Statistics Textbook. A.S.C. Ehrenberg. Wiley, 1982, xviii + 305 pp, $51.95. [ISBN: 0-471-10134-6] Nonstandard introduction, divided into six parts: statistical data, frequency distributions (and probability), sampling (and inference), relationships, communicating data, and empirical generalization. "Plays down...statistical inference from samples and tests of significance." Discusses several advanced topics, but the treatment is terse. Note price. RSK

Statistics, T(17-18: 1, 2), P. The Foundations of Multivariate Analysis: A Unified Approach by Means of Projection Onto Linear Subspaces. Kei Takeuchi, Haruo Yanai, Bishwa Nath Mukherjee. Halsted Pr, 1982, xi + 458 pp, $24.95. [ISBN: 0-85226-964-1] Emphasizes algebraic derivations and geometrical interpretations rather than statistical distributions. Techniques covered include regression analysis, analysis of variance and covariance, principal component analysis, canonical correlation analysis and discriminant analysis, factor analysis, analysis of categorical data, theory of distance and its application to classification problems, and analysis of covariance structures. RSK

Statistics, P. Proceedings of the Twenty-Seventh Conference on the Design of Experiments. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC), 1981, xvi + 511 pp, (P). Report of a conference held at North Carolina State University in Raleigh, North Carolina in October, 1981. Contains five survey lectures (including discrete distributions, exploratory data analysis, cumulative processes) and contributed papers. LAS

Statistics, S(13-18), L. Statistical Tables for the Social, Biological and Physical Sciences. F.C. Powell. Cambridge U Pr, 1982, 96 pp, $15.95; $5.95 (P). [ISBN: 0-521-24141-3; 0-521-28473-2] Tail probabilities and/or critical values for many common distributions to help carry out statistical tests, parametric and non-parametric. Also brief tables of logs and random numbers. Includes good explanations of the use of the tables. FLW

Statistics, S(15-18), P, L. Topics in Applied Multivariate Analysis. Ed: Douglas M. Hawkins. Cambridge U Pr, 1982, 362 pp, $24.95. [ISBN: 0-521-24368-8] Papers presented at a 1981 NRIMS Summer Seminar. Discriminant analysis, covariance structures, the log-linear model, scaling to reduce dimensionality, automatic interaction detection, and cluster analysis. Discusses methods that are widely applied and often little understood by their users. FLW

Statistics, S(14-18), P, L*. Comparative Statistical Inference, Second Edition. Vic Barnett. Wiley, 1982, xvi + 325 pp, $38. [ISBN: 0-471-10076-5] A thorough revision to "reflect changes of substance and attitude." The purpose is still to "develop the various principles and methods of inference and decision making to such a level that the readers can appreciate...the different approaches." (TR, First Edition, November 1974.) FLW

Statistics, T(16-18: 1), S. Design and Analysis of Experiments. M.N. Das, N.C. Giri. Halsted Pr, 1979, vii + 295 pp, $14.95. [ISBN: 0-470-26861-1] Complete and incomplete block designs, split-plot designs, factorial experiments, bio-assays, response surfaces, weighing designs, analysis of covariances. Uses partial derivatives but little matrix algebra. FLW

Computer Literacy, S. Word Processing and Text Editing. John Zarrella. Microcomputer Applications, 1980, 148 pp, $8.95 (P). [ISBN: 0-935230-01-7] Elementary introduction to the vocabulary of word processing. LAS

Computer Programming, L? Practical Pascal Programs. Gregory Davidson. Osborne/McGraw-Hill, 1982, ix + 205 pp, $15.99 (P). [ISBN: 0-931988-74-8] A collection of programs translated into Pascal from Basic. The programs represent a variety of problem areas including finance, management, and statistics. AO

Computer Programming, S(13), P. Your ATARI Computer: A Guide to ATARI 400/800 Computers. Lon Poole, Martin McNiff, Steven Cook. Osborne/McGraw-Hill, 1982, vi + 458 pp, $15 (P). [ISBN: 0-931988-65-9] A guide for users of ATARI computers. Includes a detailed introduction to programming in ATARI Basic, an introduction to ATARI hardware and its use and a complete reference manual. CEC

Computer Programming, S*, L. Problem Solving and Computer Programming. Peter Grogono, Sharon H. Nelson. Addison-Wesley, 1982, xvi + 284 pp, $14.95 (P). [ISBN: 0-201-02460-8] In two distinct parts: four chapters on general methods of solving problems (illustrated with many classic problems) followed by six on specific principles of good programming (illustrated with examples in Pascal). Intended for students with some programming experience, the book aims to help students separate the task of solving a problem from the task of writing a program, so their thinking does not become dominated by the structures of a particular language. LAS

Computer Programming, T(13: 1). A First Course in Computer Programming Using Pascal. Arthur M. Keller. McGraw-Hill, 1982, xiii + 306 pp, $14.95 (P). [ISBN: 0-07-033508-7] A thorough introduction to programming in the spirit of CS 1 of Curriculum ´78, emphasizing procedures as the first control structure. The elegance of the text is matched by the elegance of its appearance: it was prepared and typeset at Stanford in TEX, and shows off some of TEX´s multiple font features. LAS

Computer Science, P. Multicomputers and Image Processing: Algorithms and Programs. Ed: Kendall Preston, Jr., Leonard Uhr. Academic Pr, 1982, xx + 470 pp, $34. [ISBN: 0-12-564480-9] Thirty-four papers from a meeting at Madison, Wisconsin in May, 1981. Architecture, languages, algorithms; some general articles and some on special systems for image processing. RWN

Computer Science, P. The Handbook of Random Number Generation and Testing with TESTRAND Computer Code. Edward J. Dudewicz, Thomas G. Ralley. American Sci Pr, 1981, xi + 634 pp, $95 (P). [ISBN: 0-935950-01-X] Mainly listings and results of the TESTRAND package which includes 20 generators and 15 tests. Some history and general discussion. Should be considered by organizations where good random number generation is critical. RWN

Applications, S(17-18), P, L*. New Directions in Applied Mathematics. Ed: Peter J. Hilton, Gail S. Young. Springer-Verlag, 1982, ix + 163 pp, $24. [ISBN: 0-387-90604-5] Seven lectures on modern topics in applied mathematics--combinatorics, control theory, system theory, operations research, Lie groups, nonlinear analysis, turbulence--employing methods as diverse as differential and algebraic topology, theory of sets, catastrophe theory, fibre bundles, singularity theory and functional analysis. Papers presented at Case Western Reserve in April 1980 in honor of the Case centennial. Concludes with an essay on the nature and pedagogy of applied mathematics by Peter Hilton. LAS

Applications, L. Catastrophe Theory and Applications. Ed: D.K. Sinha. Halsted Pr, 1981, xii + 158 pp, $15.95. [ISBN: 0-470-27303-8] A coherent exposition of catastrophe theory, edited from notes taken at a 1979 seminar held at Jadavpur University, Calcutta. Major lectures by E.C. Zeeman are followed by various applications and a comprehensive bibliography. LAS

Applications (Artificial Intelligence), T(16-17: 1, 2), P, L. Principles of Artificial Intelligence. Nils J. Nilsson. Tioga Pub, 1980, xv + 476 pp, $30. [ISBN: 0-935382-01-1] Eminently readable introduction to fundamental ideas of artificial intelligence. Organization is based on general computational concepts involving data structures and search control strategies, rather than by areas of application. Stresses role of production systems and predicate calculus theorem proving. Exposition is simple (hence fun and easy to understand), but with excellent guidance to the literature for more details. Not enough exercises. GHM

Applications (Behavioral Science), T(15-17), L. Measurement Theory for the Behavioral Sciences. Edwin E. Ghiselli, John P. Campbell, Sheldon Zedeck. WH Freeman, 1981, xv + 494 pp, $30.75; $19.95 (P). [ISBN: 0-7167-1048-X; 0-7167-1252-0] Extensive revision of first author´s Theory of Psychological Measurement. Appeals to broader range of backgrounds and interests. Assumes introductory psychology and descriptive statistics. Equal emphasis on technical features and assumptions of various models. Thorough presentation covers such topics as correlation, regression, reliability and validity of measurement, psychological scaling and testing. JRG

Applications (Biology), P. Biomathematics in 1980. Ed: Luigi Ricciardi, Alwyn Scott. Math. Stud., V. 58. Elsevier North-Holland, 1982, xiv + 297 pp, $55.75 (P). [ISBN: 0-444-86355-9] Papers from an April 1980 international workshop at the University of Salerno, on a wide range of topics: neuroscience, population dynamics, self-organizing systems, cardiovascular modelling. LAS

Applications (Cybernetics). Progress in Cybernetics and Systems Research, V. VIII-XI. Ed: Robert Trappl, George J. Klir, Franz R. Pichler. Hemisphere Pub, 1982, $88 each. V. VIII: General Systems Methodology, Mathematical Systems Theory, Fuzzy Sets. xxiii + 529 pp [ISBN: 0-89116-237-2]; V. IX: Cybernetics in Biology and Medicine, Cybernetics in Cognition and Learning, Health Care Systems. xii + 532 pp [ISBN: 0-89116-238-0]; V. X: Structure and Dynamics of Socioeconomic Systems, Cybernetics in Organization and Management, Engineering Systems Methodology, Systems Research on Science and Technology. xiii + 562 pp [ISBN: 0-89116-239-9]; V. XI: Data Base Design, International Information

Systems, Semiotic Systems, Artificial Intelligence, Cybernetics and Philosophy, Special Aspects. xv + 601 pp. [ISBN: 0-89116-240-2] These volumes represent part of the proceedings of the Fifth European meeting on cybernetics and systems research that took place at the University of Vienna, April 8-11, 1980. Earlier volumes represent proceedings of other European conferences on the same topics. JAS

Applications (Economics), T(15-17: 1), S, P, L. Economics for Mathematicians. J.W.S. Cassels. London Math. Soc. Lect. Note Ser., No. 62. Cambridge U Pr, 1981, xi + 145 pp, $14.95 (P). [ISBN: 0-521-28614-X] Concise notes expressing classical economic theories in standard mathematical language (linear algebra, multivariable calculus, convex sets). A good crash course in economic theories and terminology, but without the real-world examples of standard treatments. LAS

Applications (Mechanics of Solids), T?(18: 1), P. Theory of Thin Elastic Shells. M. Dikmen. Pitman Pub, 1982, xii + 364 pp, $59.95. [ISBN: 0-273-08431-3] The objects of study are deformable, simply connected, 3-dimensional bodies which are "thin" in that they admit a tolerable 2-dimensional analysis. This comprehensive monograph develops several mathematical theories treating shells. The level is advanced, but there are several mathematical supplements. PZ

Applications (Physics), T(14-18: 1, 2), S, L. An Introduction to Tensor Calculus, Relativity and Cosmology, Third Edition. D.F. Lawden. Wiley, 1982, xiii + 205 pp, $35.95. [ISBN: 0-471-10082-X] First four chapters are devoted to the special theory of relativity including Lorentz transformations, orthogonal transformations, Cartesian tensors, mechanics, electrodynamics. Remaining three chapters focus on general theory of relativity: general tensor calculus, Riemannian space, cosmology. Chapter exercises. References, bibliography, index. RJA

Applications (Physics), P. Complex Geometry in Mathematical Physics. R.O. Wells, Jr. Pr U Montreal, 1982, 168 pp, $14 (P). [ISBN: 2-7606-0559-8] An introduction to some of the work done in mathematical physics in the last ten years. The underlying machinery is complex twistor geometry growing out of the ideas of Roger Penrose; the main thrust is the description of solutions of certain problems in field theory. JAS

Applications (Social Science), P. Cultural Transmission and Evolution: A Quantitative Approach. L.L. Cavalli-Sforza, M.W. Feldman. Mono. in Population Biology, V. 16. Princeton U Pr, 1981, xiv + 388 pp, $25; $10.50 (P). [ISBN: 0-691-08280-4; 0-691-08283-9] An extension of many of the quantitative methods developed in population genetics to cultural evolution of learned traits. This volume treats vertical (parent to child) and horizontal (older teachers to younger students) transmission of dichotomous traits, multiple state traits, and continuous traits. Mathematical techniques are limited to elementary probability and calculus--to avoid limiting the audience for the volume. LAS

Applications (Social Science), P, L.** Watergate Games: Strategies, Choices, Outcomes. Douglas Muzzio. New York U Pr, 1982, xii + 205 pp, $17.95. [ISBN: 0-8147-5384-1] Attempt to "explain" the major developments of the Watergate investigation (but not the infamous break-in itself) by casting them in the language of game theory with various goals, strategies and payoff matrices. Author contends that the major "players" in these Watergate "games" acted quite "rationally" when measured by the optimal strategies suggested by game theory. GHM

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Ohio Section

The fall meeting of the Ohio Section was held October 22-23, 1982 at Youngstown State University with 130 individuals in attendance. The meeting was a joint meeting with SIAM.

Invited Addresses:

"A Sampling of Applied Combinatorial Problems," by David P. Roselle, Secretary of MAA, Dean of Research and Graduate Studies, Virginia Polytechnic Institute and State University.

Systems, Semiotic Systems, Artificial Intelligence, Cybernetics and Philosophy, Special Aspects. xv + 601 pp. [ISBN: 0-89116-240-2] These volumes represent part of the proceedings of the Fifth European meeting on cybernetics and systems research that took place at the University of Vienna, April 8-11, 1980. Earlier volumes represent proceedings of other European conferences on the same topics. JAS

Applications (Economics), T(15-17: 1), S, P, L. Economics for Mathematicians. J.W.S. Cassels. London Math. Soc. Lect. Note Ser., No. 62. Cambridge U Pr, 1981, xi + 145 pp, $14.95 (P). [ISBN: 0-521-28614-X] Concise notes expressing classical economic theories in standard mathematical language (linear algebra, multivariable calculus, convex sets). A good crash course in economic theories and terminology, but without the real-world examples of standard treatments. LAS

Applications (Mechanics of Solids), T?(18: 1), P. Theory of Thin Elastic Shells. M. Dikmen. Pitman Pub, 1982, xii + 364 pp, $59.95. [ISBN: 0-273-08431-3] The objects of study are deformable, simply connected, 3-dimensional bodies which are "thin" in that they admit a tolerable 2-dimensional analysis. This comprehensive monograph develops several mathematical theories treating shells. The level is advanced, but there are several mathematical supplements. PZ

Applications (Physics), T(14-18: 1, 2), S, L. An Introduction to Tensor Calculus, Relativity and Cosmology, Third Edition. D.F. Lawden. Wiley, 1982, xiii + 205 pp, $35.95. [ISBN: 0-471-10082-X] First four chapters are devoted to the special theory of relativity including Lorentz transformations, orthogonal transformations, Cartesian tensors, mechanics, electrodynamics. Remaining three chapters focus on general theory of relativity: general tensor calculus, Riemannian space, cosmology. Chapter exercises. References, bibliography, index. RJA

Applications (Physics), P. Complex Geometry in Mathematical Physics. R.O. Wells, Jr. Pr U Montreal, 1982, 168 pp, $14 (P). [ISBN: 2-7606-0559-8] An introduction to some of the work done in mathematical physics in the last ten years. The underlying machinery is complex twistor geometry growing out of the ideas of Roger Penrose; the main thrust is the description of solutions of certain problems in field theory. JAS

Applications (Social Science), P. Cultural Transmission and Evolution: A Quantitative Approach. L.L. Cavalli-Sforza, M.W. Feldman. Mono. in Population Biology, V. 16. Princeton U Pr, 1981, xiv + 388 pp, $25; $10.50 (P). [ISBN: 0-691-08280-4; 0-691-08283-9] An extension of many of the quantitative methods developed in population genetics to cultural evolution of learned traits. This volume treats vertical (parent to child) and horizontal (older teachers to younger students) transmission of dichotomous traits, multiple state traits, and continuous traits. Mathematical techniques are limited to elementary probability and calculus--to avoid limiting the audience for the volume. LAS

Applications (Social Science), P, L.** Watergate Games: Strategies, Choices, Outcomes. Douglas Muzzio. New York U Pr, 1982, xii + 205 pp, $17.95. [ISBN: 0-8147-5384-1] Attempt to "explain" the major developments of the Watergate investigation (but not the infamous break-in itself) by casting them in the language of game theory with various goals, strategies and payoff matrices. Author contends that the major "players" in these Watergate "games" acted quite "rationally" when measured by the optimal strategies suggested by game theory. GHM

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Ohio Section

The fall meeting of the Ohio Section was held October 22-23, 1982 at Youngstown State University with 130 individuals in attendance. The meeting was a joint meeting with SIAM.

Invited Addresses:

"A Sampling of Applied Combinatorial Problems," by David P. Roselle, Secretary of MAA, Dean of Research and Graduate Studies, Virginia Polytechnic Institute and State University.

"The Process of Revenue Estimating for State Budget Purposes," by Samuel Nemer, Office of Budget and Management, State of Ohio.
"Results of Testing College-intending High School Juniors," by Bert K. Waits, Director Early Mathematics Placement Testing Program," Ohio State University.
"What is Applied Mathematics? What is Applied Mathematics Education?" by James M. Greenberg, Ohio State University; formerly NSF Program Director for Applied Mathematics.

Panel Discussions:

Discussion of Ohio Section Survey of Recent Mathematical Sciences Graduates, by Tom Price, University of Akron, Moderator.
Report on Retraining for Computer Science, by Zaven Karian, Denison University, Moderator.
"Nurturing Young Talent," by Robert Dieffenbach, Miami University.

Contributed Papers:

"Kind-of-5 Plus Sort-of-8 is Imprecisely Close to Sort-of-5 Plus Kind-of-8," by Steven Rodabaugh, Youngstown State University.
"A Non-commutative Pettis Theorem," by Nazanin Azarnia, Miami University, Hamilton.
"Mathematics and Art," by Gus Mavrigan, Youngstown State University.
"Almost Periodic Functions on the Real Line," by David Colella, Youngstown State University.
"Poles and Zeros of Matrices of Functions," by Bostwick Wyman, Ohio State University.
"Studying Curves at L_∞," by Kenneth Cummins, Kent State University.
"A Metrization Theorem, or Can This Dog Be Saved?" by Albert Klein, Youngstown State University.
"Independent Partitions and the Unimodality Conjecture," by Thomas A. Dowling, Ohio State University.
"Difference Sets," by T. Arasu, Ohio State University.
"Association Schemes," by Dijen K. Ray-Chaudhuri, Ohio State University.

Indiana Section

The fall meeting of the Indiana Section was held at Wabash College on Saturday, October 16, 1982 with 50 members present.

Student Paper:

"The German Enigma," by Brian Wade and Tony Kirk, Rose-Hulman Institute of Technology.

Contributed Presentations:

"Some Applications of Logic to Algebra," by Leonard Lipshitz, Purdue University.
"Baire Category: A Proof in Search of a Theorem," by William Dunham, Hanover College.
"Some Comments on Articulation," by Billy Rhodes, Indiana University.
"Numerical Solution of Two-point Boundary-value Problems: A Survey," by James Daniel, University of Texas.

North Central Section

The North Central Section held its fall meeting at the University of Minnesota at Duluth on October 29-30, 1982.

Invited Addresses:

"Know-How in Mathematics," by Loren C. Larson, St. Olaf College.
"Applications of Graph Theory to Linear Algebra," by Michael Doob, University of Manitoba.

Short Presentations:

* "Another Proof that A_5 is Simple," by Joseph Gallian, University of Minnesota at Duluth.
* "Growth Functions of Groups," by Max Benson, University of Minnesota at Duluth.
* "Is Time Real, If You Have to Tell It (A New Concept of Time is Needed)," by Robert L. Jacobsen.
"Fixed Points for Set-valued Maps," by K.L. Singh, University of Minnesota at Duluth.
* "Integer-sided Triangles Whose Ratio of Base to Altitude is an Integer," by Gerald E. Bergum, South Dakota State University.
"LOGO for Learning Mathematics," by Roger Kirchner, Carleton College.
"Iterative Methods for Solving Systems of Equations," by Larry F. Bennett, South Dakota State University.
* "An Illustrated History of Remainders," by Don Mattson and James Hatzenbuhler, Moorhead State University.
* "Mathletes in Minnesota," by Wayne Roberts, Macalester College.
"Porosity, A Nice Concept in Exceptional Sets," by Ted Vessey, St. Olaf College.

Mathematics Appreciation Courses: The Report of a CUPM Panel

Bibliography and Reference List

A report called "Mathematics Appreciation Courses" appears in the Teaching of Mathematics section of this issue of the Monthly. This report in its original form contained an extensive list of references (films, supplies, books) for teachers of such courses. The following pages contain this list of references.

Films

A Non-Euclidean Universe. (1978; 25 Min; Color). University Media.

A Time for Change--The Calculus. (1975; 25 Min; Color) University Media.

Accidental Nuclear War. (1976; 8 Min; Color) Pictura Film.

Adventures in Perception. (1973; 22 Min; Color) BFA Educational Media. (Reviews: Amer. Math. Monthly 84 (1977) 582.)

An Historical Introduction to Algebra. Modern Film Rentals.

An Introduction to Feedback. (1960; 11 Min; Color) Encyclopedia Britannica Educational Corporation.

Anatomy of Analogy. (25 Min; BW) Open University.

Area and Pi. (1969; 10 Min; Color) Modern Film Rentals.

Auto Insurance. (1976; 8 Min; Color) Pictura Film.

Caroms. (1971; 9 1/2 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 417; Math Teacher 66 (1973) 51.)

Central Perspectivities. (1971; 13 1/2 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 419; Math. Teacher (1972) 733.)

Central Similarities. (1966; 10 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 418; Math. Teacher (1972) 643-644.)

Challenge in the Classroom. (1966; 55 Min; Color) Modern Film Rentals.

Circle Circus. (1979; 7 Min; Color) International Film Bureau.

Common Generation of Conics. (4 Min; Color). Educational Solutions.

Complex Numbers. (1978; 25 Min; Color). University Media.

Computer Perspective. (1972; 8 Min; Color) Pyramid Films.

Congruent Triangles. (1978; 7 Min; Color). International Film Bureau.

Conic Sections. (1968; 11 Min; Color) BFA Educational Media.

Conics. (1979; 10 Min; Color) Wards Modern Learning Aids.

Conics. (1978; 25 Min; Color). University Media.

Constructing an Algorithm. (25 Min; BW) Open University.

Cosmic Zoom. McGraw-Hill Films.

Curves. (1968; 17 Min; Color) A.I.M.S. (Reviews: Amer. Math. Monthly 83 (1976) 71-72; Math. Teacher 64 (1971) 525.)

Curves of Constant Width. (1971; 16 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 78 (1971) 539; Math. Teacher 65 (1972) 234.)

Cycloidal Curves or Tales From the Wanklenberg Woods. (1974; 22 Min; Color) Modern Film Rentals.

Dance Squared. (1963; 4 Min; Color) International Film Bureau. Review: Math. Teacher 64 (1971) 627.)

Dihedral Kaleidoscopes. (1966; 13 Min; Color) International Film Bureau. (Review: Math. Teacher 66 (1973) 51.)

Dimension. (1970; 13 Min; Color) A.I.M.S. (Reviews: Amer. Math. Monthly 83 (1976) 71-72; Math. Teacher 64 (1971) 525.)

Donald in Mathmagicland. (1960; 26 Min; Color) Walt Disney Educational Media Company.

Dr. Posin's Giants: Isaac Newton. Indiana University Audiovisual Library.

Dragon Fold...And Other Ways to Fill Space. (1979; 7 1/2 Min; Color) International Film Bureau.

Equidecomposable Polygons. (10 1/2 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 687-688; Math. Teacher 65 (1972) 734.)

Errors That Die. (25 Min; BW) Open University.

Flatland. (1965; 12 Min; Color) McGraw-Hill Films. (Review: Math. Teacher 64 (1971) 44-45.)

Functions and Graphs. (1978; 25 Min; Color) University Media.

Geodesic Domes: Math Raises the Roof. (1979; 20 Min; Color) David Nulsen Enterprises.

Geometric Vectors--Addition. (1971; 17 Min; Color) International Film Bureau. (Review: Amer. Math. Monthly 82 (1975) 420.)

Geometry: Inductive and Deductive Reasoning. (1962; 12 1/2 Min; Color) Coronet Films.

Good for What? (25 Min; BW) Open University.

Gottingen and New York. (1966; 43 Min; Color) Modern Film Rentals.

How Far is Around? (1979; 7 1/2 Min; Color) International Film Bureau.

Inferential Statistics, Part I: Sampling and Estimation. (1977; 19 Min; Color) Media Guild.

Inferential Statistics, Part II: Hypothesis Testing. (1977; 25 Min; Color). Media Guild.

Infinity. (1972; 17 Min; Color) A.I.M.S. (Review: Amer. Math. Monthly 83 (1976) 71-72.)

Inversion. (12 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 83 (1976) 71; Math. Teacher (1972) 644.)

Isaac Newton. (1959; 13 1/2 Min; Color) Coronet Films.

Isn't That the Limit! (1980; 17 Min; Color) David Nulsen Enterprises.

Isometries. (1967; 26 Min; Color) International Film Bureau. (Review: Math. Teacher 66 (1973) 51-52.)

Iteration and Convergence. (1978; 25 Min; Color) University Media.

John von Neumann, A Documentary. (1966; 63 Min; BW) Modern Film Rentals. (Review: Amer. Math. Monthly 75 (1968) 435.)

Journey to the Center of a Triangle. (1977; 8 1/2 Min; Color) International Film Bureau.

Let Us Teach Guessing. (1966; 61 Min; Color) Modern Film Rentals. (Review: Amer. Math. Monthly 75 (1968) 219.)

Limit Curves and Curves of Infinite Length. (1979; 14 Min; Silent; Color) International Film Bureau.

Limit Surfaces and Space Filling Curves. (1979; 10 1/2 Min; Silent; Color) International Film Bureau.

Limits. (25 Min; BW) Open University.

Linear Programming. (1969; 9 Min; Color) Macmillan Films.

Look Again. (1970; 15 Min; Color) A.I.M.S. (Reviews: Amer. Math. Monthly 83 (1976) 71-72; Math. Teacher 64 (1971) 525.)

Love Song. (1976; 11 Min; Color) Pictura Film.

Mathematical Curves. (1977; 10 Min; Color) Churchill Films.

Mathematical Induction. (1960; 62 Min; Color) Modern Film Rentals.

Mathematical Induction. (1978; 25 Min; Color) University Media.

Mathematical Peep Show. (1961; 11 Min; Color) Encyclopedia Britannica Educational Corporation. (Review: Math. Teacher 64 (1971) 625.)

Mathematician and the River. (1959; 19 Min; Color) No distributor.

Mathematics of the Honeycomb. (1964; 13 Min; Color) Moody Institute of Science. (Review: Math. Teacher 64 (1971) 334.)

Matrices. (9 Min; Color) Macmillan Films.

Matrioska. Indiana University Audiovisual Library.

Maurits Escher, Painter of Fantasies. (1970; 26 1/2 Min; Color) Coronet Films. (Review: Amer. Math. Monthly 83 (1976) 495.)

Mean, Median, Mode. McGraw-Hill Films.

Modelling Drug Therapy. (1978; 25 Min; Color). University Media.

Modelling Pollution. (1978; 25 Min; Color). University Media.

Modelling Surveys. (1978; 25 Min; Color). University Media.

Modmath. (14 1/2 Min; Color) International Film Bureau.

Mr. Simplex Saves the Aspidistra. (1966; 33 Min; Color) Modern Film Rentals.

Networks and Matrices. (1978; 25 Min; Color) University Media.

New Worlds From Old. (1975; 25 Min; Color) University Media.

Newton's Equal Areas. (1968; 8 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 79 (1972) 1054; Math. Teacher 63 (1970) 449.)

Nim and Other Oriented Graph Games. (1966; 63 Min; BW) Modern Film Rentals.

Notes on a Triangle. International Film Bureau. (Review: Math. Teacher 63 (1970) 363.)

Numbers Now and Then. (1975; 25 Min; Color) University Media.

Orthogonal Projection. (1965; 13 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 419-420; Math. Teacher (1972) 643.)

Paradox Box. Scientific American.

Pits, Peaks, and Passes (Part 1). (1966; 48 Min; Color) Modern Film Rentals.

Plateau's Problem. A Film by Sr. Rita Ehrmann.

Points of View: Perspective and Projection. (1975; 25 Min; Color) University Media.

Possibly So, Pythagoras. (1973; 14 Min; Color). International Film Bureau. (Review: Math. Teacher 64 (1971) 626.)

Powers of Ten. (1978; 9 Min; Color) Pyramid Films. (Review: Math. Teacher (1979) 388.)

Predicting at Random. (1966; 43 Min; Color) Modern Film Rentals.

Probability. (12 Min; Color) McGraw-Hill Films.

Professor George Pólya and Students, Parts I and II. (1972; 60 Min; Color) University Media.

Professor George Pólya Talks to Professor Maxim Bruckbeimer. (1972; 60 Min; Color) University Media.

Projective Generation of Conics. (16 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 82 (1975) 538-539; Math. Teacher 66 (1973) 51.)

Quaternions: A Herald of Modern Algebra. (1975; 25 Min; Color) University Media.

Rational Numbers and the Square Root of 2. (1978; 25 Min; Color) University Media.

Regular Homotopies in the Plane: Part I: (1975; 14 Min; Color); Part II: (1975; 18 1/2 Min; Color) International Film Bureau. (Review: Amer. Math. Monthly 85 (1978) 212.

Root Two: Geometry or Arithmetic? (1975; 25 Min; Color) University Media.

Sampling. (25 Min; BW) Open University.

Sets, Crows, and Infinity. (12 Min; Color) BFA Educational Media.

Shaking the Foundations. (1975; 25 Min; Color) University Media.

Shapes of the Future: Some Unsolved Problems in Geometry: Part I: Two Dimensions (1975; 22 Min; Color); Part II: Three Dimensions (1970; 21 Min; Color) Modern Film Rentals. (Review: Amer.

Math. Monthly 79 (1972) 1052-1053.)

Sierpinski's Curve Fills Space. (1979; 4 1/2 Min; Color) International Film Bureau.

Similar Triangles. (1976; 7 1/2 Min; Color) International Film Bureau.

Space Filling Curves. (1975; 25 1/2 Min; Color) International Film Bureau. (Review: Math. Teacher 69 (1976) 164-165.)

Sphere Eversions. (1979; 7 1/2 Min; Silent; Color) International Film Bureau.

Spheres. International Film Bureau.

Statistics At A Glance. (1972; 28 Min; Color) Media Guild. (Review: Amer. Math. Monthly 82 (1975) 312.)

Statistics and Probability I. (15 Min; BW) Open University.

Statistics and Probability II. (25 Min; BW) Open University.

Statistics and Probability III. (25 Min; BW) Open University.

Symbols, Equations and the Computer. (1978; 25 Min; Color) University Media.

Symmetries of the Cube. (1971; 13 1/2 Min; Color) International Film Bureau. (Review: Math. Teacher 65 (1972) 733.)

The Algebra of the Unknown. (1975; 25 Min; Color) University Media.

The Binomial Theorem. (1978; 25 Min; BW) University Media.

The Butterfly Catastrophe. (1979; 4 1/2 Min; Silent; Color) International Film Bureau.

The Delian Problem. (1975; 25 Min; Color) University Media.

The Dot and the Line. Indiana University Audiovisual Library.

The Geometry Euclid Didn't Know. (1979; 16 Min; Color) David Nulsen Enterprises. (Reviews: Amer. Math. Monthly 86 (1979) 600; Math. Teacher (1979) 300.)

The Great Art--Solving Equations. (1975; 25 Min; Color) University Media.

The Hypercube: Projections and Slicing. (1978; 12 Min; Color) Banchoff-Strauss Productions.

The Kakeya Problem. (1962; 60 Min; Color) Modern Film Rentals.

The Majestic Clockwork. (1974; 52 Min; Color) Time Life Multimedia.

The Marriage Theorem, Parts I & II. (1974; 46 Min. and 47 Min; BW) Modern Film Rentals.

The Music of the Spheres. (1974; 52 Min; Color) Time Life Multimedia.

The Nature of Digital Computing. (25 Min; BW) University Media.

The Perfection of Matter. McGraw-Hill Films.

The Search for Solid Ground. (1963; 62 Min; BW) Modern Film Rentals.

The Seven Bridges of Königsberg. (1965; 4 Min; Color) International Film Bureau.

The Structure of a Computer. (25 Min; BW) Open University.

Topology. (1972; 9 Min; Color) Macmillan Films.

Topology. (1966; 30 Min; BW) Modern Film Rentals. (Review: Amer. Math. Monthly 75 (1968) 790.)

Topology: Some Historical Concepts. (21 3/4 Min; Color) Richard Cline Film Productions.

Trio for Three Angles. (8 Min; Color) International Film Bureau.

Turning a Sphere Inside Out. (1976; 23 Min; Color) International Film Bureau. (Reviews: Amer. Math. Monthly 86 (1979) 511-512; Math. Teacher 70 (1977) 55.)

View From the People Wall. (1964; 14 Min; Color) Encyclopedia Britannica Educational Corporation.

Weather by the Numbers. University of Indiana Audiovisual Library.

What is a Limit? (25 Min; BW) Open University.

What is a Set? Part I & II. (1967; 15 Min; Color) Modern Film Rentals. (Review: Amer. Math. Monthly 75 (1968) 324.)

What is Mathematics and How Do We Teach it? (1966; 60 Min; BW) Modern Film Rentals.

Zooms on Self-Similar Figures. (1979; 8 Min; Color) International Film Bureau.

Addresses of Distributors:

A.I.M.S., 626 Justin Avenue, Glendale, CA 91201.

International Film Bureau, 332 South Michigan Avenue, Chicago, IL 60604.

McGraw-Hill Films, McGraw-Hill Book Company, 330 West 42nd Street, New York, NY 10036.

Modern Film Rentals, 2323 New Hyde Park Road, New Hyde Park, NY 11040.

Moody Institute of Science, 12000 East Washington Boulevard, Whittier, CA 90606.

Indiana University Audiovisual Library, Bloomington, IN 47401.

Ward's Modern Learning Aids Division, P.O. Box 1712, Rochester, NY 14603.

University Media, 118 South Acacia, Box 881, Solana Beach, CA 92075.

Pyramid Films, P.O. Box 1048, Santa Monica, CA 90406.

Educational Solutions, Inc., 80 Fifth Avenue, New York, NY 10011.

Banchoff-Strauss Productions, Inc., P.O. Box 2430, East Side Station, Providence, RI 02906.

Macmillan Films, Inc., 34 MacQuesten Parkway South, Mt. Vernon, NY 10550.

Encyclopedia Britannica Educational Corporation, 425 North Michigan Avenue, Chicago, IL 60611.

David Nulsen Enterprises, 3211 Pico Boulevard, Santa Monica, CA 90405.

The Media Guild, 118 South Acacia, Box 881, Solana Beach, CA 92075.

Coronet Films, 65 East South Water Street, Chicago, IL 60601.

Pictura Film Co., 111 8th Avenue, New York, NY 10011.

BFA Educational Media, 2211 Michigan Avenue, P.O. Box 1795, Santa Monica, CA 90406.

Time-Life Multimedia, Time and Life Building, New

York, NY 10020.

Churchill Films, 662 North Robertson Boulevard, Los Angeles, CA 90069.

Walt Disney Educational Media Company, 500 South Buena Vista Street, Burbank, CA 91521.

Contemporary-McGraw Hill Films, 1221 Avenue of the Americas, New York, NY 10020.

Suppliers of Classroom Aids

The Math Group, Inc., 396 East 79th Street, Minneapolis, MN 55420.
Unique puzzles and card games. Designed for elementary school enrichment, many are flexible enough to be of interest to adults as well.

Lano Company, 9001 Gross Road, Dexter, MI 48130. (313-426-4860)
Mathematical visual aids (solids, transparencies, graphing aids) together with various games and probability models.

Math Shop, Inc., 5 Bridge Street, Watertown, MA 02172.
A full range of curriculum enrichment material for elementary, junior and senior high school mathematics. Includes several games and puzzles of use in introductory college mathematics.

International Film Bureau, Inc., 332 South Michigan Avenue, Chicago, IL 60604.
The major American distributor of mathematics films. The current list includes nearly 50 films at the high school and college level, each of which may be either purchased or rented.

Yoder Instruments, East Palestine, OH 44413. (216-426-3612; 216-426-9580).
Two unique geometric construction sets for plane and solid geometry.

Cuisenaire Co. of America, Inc., 12 Church Street, New Rochelle, NY 10805.
Sensory apparatus designed for elementary schools, some of which (e.g., geoboards, polyhedral structures) would be suitable to courses in mathematics appreciation.

LaPine Scientific Company, Department B43, 6005 South Knox Avenue, Chicago, IL 60629; 373 Chestnut Street, Norwood, NJ 07648; 920 Parker Street, Berkeley, CA 94710; Box 95, Postal Station U., Toronto, Canada M8Z 5M4.
An extensive offering of models, teaching aids, games, and audio-visual materials for elementary, high school and beginning college mathematics. Mathematics catalogue exceeds 100 pages.

Geyer Instructional Aids Co., Inc., P.O. Box 7306, Fort Wayne, IN 46807.
A large collection of models, games, books and classroom aids for high school and elementary college courses.

Creative Publications, 3977 East Bayshore Road, P.O. Box 10328, Palo Alto, CA 94303. (415-968-3977).
An attractive 100-page catalogue of mathematical books, models, games, posters, construction sets, and puzzles for all grade levels. The premier source for mathematics enrichment material.

Inquiry Audio-Visuals, 1754 West Farragut Avenue, Chicago, IL 60640.
Filmstrips and transparencies for high school algebra topics.

Educational Audio Visual, Inc., Pleasantville, NY 10570.
Transparencies and games for algebra, geometry calculus and statistics.

W.H. Freeman and Company, 660 Market Street, San Francisco, CA 94104.
Offprints of Scientific American articles.

David Nulsen Enterprises, 3211 Pico Boulevard, Santa Monica, CA 90405.
Three 15-minute, 16mm color films ("Curves," "Dimension," "Look Again") for rent or purchase.

Popular Science Audio-Visuals, Inc., 5235 Ravenswood Avenue, Chicago, IL 60640.
Filmstrips and overhead transparencies, principally in high school mathematics (geometry, algebra, elementary functions).

Time Life Films, 43 West 16th Street, New York, N 10011. (212-691-2930).
Numerous BBC-produced 20-minute 16mm B & W films on topics ranging from inequalities to matrices.

Edmund Scientific Company, 1985 Edscorp Building, Barrington, NJ 08007. (609-547-3488).

References

Survey Monographs

Boehm, George A.W. The New World of Mathematics. The Dial Press, New York, 1959.

Courant, Richard and Robbins, Herbert. What is Mathematics? Oxford University Press, New York, 1941.

Dantzig, Tobias. Number, the Language of Science 4th ed. Free Press, New York, 1967.

Garding, Lars. Encounter with Mathematics. Springer-Verlag, New York, 1977.

Herstein, I.N. and Kaplansky, I. Matters Mathematical. Harper and Row, New York, 1974.

Khurgin, Ya. Did You Say Mathematics? MIR Publishers, Moscow, Russia, 1974.

Kline, Morris. Mathematics: The Loss of Certainty. Oxford University Press, New York, 1980.

Pedoe, Daniel. The Gentle Art of Mathematics. Macmillan, Riverside, New Jersey, 1958, 1963.

Rademacher, Hans and Toeplitz, Otto. The Enjoyment of Mathematics. Princeton University Press, Princeton, New Jersey, 1957.

Sawyer, W.W. Introducing Mathematics, 4 vols. Penguin Books, New York, 1964-70.

Singh, Jagjit. Great Ideas of Modern Mathematics, Their Nature and Use. Dover, New York, 1959; Hutchinson and Company, London, England, 1972.

Stein, Sherman K. Mathematics, The Man-made Universe: An Introduction to the Spirit of Mathematics, Third Edition. W.H. Freeman, San Francisco, California, 1976.

Steinhaus, Hugo. Mathematical Snapshots, 2nd ed. Oxford University Press, New York, 1969.

Stewart, Ian. Concepts of Modern Mathematics. Penguin Books, New York, 1975.

Whitehead, Alfred North. An Introduction to

Mathematics. Oxford University Press, New York, 1958.

Collections of Essays

LeLionnais, F., ed. Great Currents of Mathematical Thought, 2 vols. Dover, New York, 1971.

Kline, Morris, ed. Mathematics in the Modern World. W.H. Freeman, San Francisco, California, 1968.

Kline, Morris, ed. Mathematics: An Introduction to Its Spirit and Use. W.H. Freeman, San Francisco, California, 1979.

Messick, David M., ed. Mathematical Thinking in Behavioral Sciences. W.H. Freeman, San Francisco, California, 1968.

National Research Committee on Support of Research in the Mathematical Sciences (COSRIMS). The Mathematical Sciences-- A Collection of Essays. MIT Press, Cambridge, Massachusetts, 1969.

Newman, James R., ed. The World of Mathematics, 4 vols. Simon and Schuster, New York, 1956-60.

Saaty, Thomas L. and Weyl, F. Joachim, eds. The Spirit and Uses of the Mathematical Sciences. McGraw-Hill, New York, 1969.

Schaaf, William L., ed. Our Mathematical Heritage, New, Revised Edition. Collier Books, New York, 1963.

Steen, Lynn Arthur, ed. Mathematics Today: Twelve Informal Essays. Springer-Verlag, New York, 1978.

Nature of Mathematics

Adler, Alfred. "Reflections--mathematics and creativity." New Yorker 47 (February 19, 1972) 39-45.

Bronowski, Jacob. "The music of the spheres." in J. Bronowski, The Ascent of Man. Little, Brown, and Company, Boston, Massachusetts, 1973, pp. 154-187.

Bronowski, Jacob. "The logic of the mind." Amer. Scientist 54 (1966) 1-14.

Bruter, C.P. Sur la nature des mathématiques. Gauthier-Villars, Paris, France, 1973.

Cartwright, Mary L. "The mathematical mind." Math. Spectrum 2 (1969-70) 37-45.

Cartwright, Mary L. "Mathematics and thinking mathematically." Amer. Math. Monthly 77 (1970) 20-28.

Davis, Philip J. and Hersh, Reuben. The Mathematical Experience. Birkhäuser, Cambridge, Massachusetts, 1981.

Fisher, Charles S. "Some social characteristics of mathematicians and their work." Amer. J. Sociology 78 (1973) 1094-1118.

Grabiner, Judith V. "Is mathematical truth time-dependent?" Amer. Math. Monthly 81 (1974) 354-365.

Hadamard, Jacques. Psychology of Invention in the Mathematical Field. Dover, New York, 1945.

Halmos, Paul R. "Mathematics as a creative art." Amer. Scientist 56 (1968) 375-389.

Hahn, Hans. "Geometry and intuition." Scientific American 190 (April 1954) 84-91, 108; also in M. Kline. Mathematics in the Modern World. W.H. Freeman, San Francisco, California, 1968, pp. 184-188, 399.

Hardy, G.H. A Mathematician's Apology. Cambridge University Press, Cambridge, Massachusetts, 1940; 1967; excerpted in J.R. Newman. The World of Mathematics, V. 4, Simon and Schuster, New York, 1956, pp. 2027-2038.

Helitzer, Florence. "A conversation with three mathematicians." University: A Princeton Quarterly 59 (Winter 1974) 1-5, 28-30.

Henkin, Leon. "Are logic and mathematics identical?" Science 138 (1962) 788-794.

Hilton, Peter J. "The art of mathematics." Univ. of Birmingham, 1960.

Iliev, L. "Mathematics as the science of models." Russian Math. Surveys 27:2 (1972) 181-189.

Jones, Landon Y., Jr. "Mathematicians: They're special." Think 40:4 (1974) 32-35.

Kapur, J.N. Thoughts on the Nature of Mathematics. Atma Ram, Delhi, India, 1973.

Lefschetz, Solomon. "The structure of mathematics." Amer. Scientist 38 (1950) 105-111.

Newman, M.H.A. "What is mathematics? New answers to an old question." Math. Gazette 43 (1959) 161-171.

Otte, Michael. Mathematiker über die Mathematik. Springer-Verlag, New York, 1974.

Poincaré, Henri. "Mathematical creation." Scientific American 179 (August 1948) 54-57; also in M. Kline. Mathematics in the Modern World. W.H. Freeman, San Francisco, California, 1968, pp. 14-17; and in J.R. Newman, The World of Mathematics, V. 4, Simon and Schuster, New York, 1956, pp. 2041-2050.

Rényi, Alfred. "A Socratic dialogue on mathematics." Canad. Math. Bull. 7 (1964) 441-462; also in A. Rényi. Dialogues on Mathematics. Holden-Day, San Francisco, California, 1967, pp. 3-25.

Stein, Sherman K. "The mathematician as an explorer." Scientific American 204 (May 1961) 148-158, 206.

Stone, Marshall H. "The revolution in mathematics." Liberal Education, 47 (1961) 304-327; also in Amer. Math. Monthly 68 (1961) 715-734.

Weidman, Donald R. "Emotional perils of mathematics." Science 149 (1965) 1048.

Weissinger, Johannes. "The characteristic features of mathematical thought." in T.L. Saaty and F.J. Weyl, The Spirit and Uses of the Mathematical Sciences. McGraw-Hill, New York, 1969, pp. 9-27.

Weyl, Hermann. "The mathematical way of thinking." Science 92 (1940) 437-446; also in Studies in the History of Science. University of Pennsylvania Press, 1941, pp. 103-123.

Weyl, Hermann. "Insight and reflection." in T.L. Saaty and F.J. Weyl. The Spirit and Uses of the Mathematical Sciences. McGraw-Hill, New York, 1969, pp. 281-301.

Wilder, Raymond L. "The role of the axiomatic method." Amer. Math. Monthly 74 (1967) 115-127; also in Math. Teaching 41 (1967) 32-40.

von Neumann, John. "The mathematician." in R.B. Heywood, The Works of the Mind. University of Chicago Press, Chicago, Illinois, 1947, pp. 180-196; also in J.R. Newman, The World of

Mathematics, V. 4, Simon and Schuster, New Yok, 1956, pp. 2053-2063.

Survey Papers

Bochner, Salomon. "Mathematics." McGraw-Hill Encyclopedia of Science and Technology 8 (1960) 175-180.

Dieudonné, Jean A. "Recent developments in mathematics." Amer. Math. Monthly 71 (1964) 239-248.

Eves, Howard W. "Mathematics." Encyclopedia Americana 18 (1976) 431-434.

Ficken, F.A. "Mathematics and the layman." Amer. Scientist 52 (1964) 419-430.

MacLane, Saunders. "Mathematical models of space." Amer. Scientist 53 (1980) 252.

Meserve, Bruce E. "New mathematics." Encyclopedia Americana 20 (1976) 202-205.

Meserve, Bruce E. "Number systems and notation." Encyclopedia Americana 20 (1976) 536f-536j.

Murray, Francis J. and Ford, Lester R. "Mathematics as a calculatory science." Encyclopaedia Britannica, 15th ed., 1974, Macropaedia V. 11, pp. 671-696.

Richards, Ian. "Impossibility." Math. Magazine 48 (1975) 249-262.

Stone, Marshall H. "The future of mathematics." J. Math. Soc. Jap. 9 (1957) 493-507.

Temple, G. "The growth of mathematics." Math. Gazette 41 (1957) 161-168.

Weil, André. "The future of mathematics." Amer. Math. Monthly 57 (1950) 295-306; also in F. LeLionnais. Great Currents of Mathematical Thought, V. 1. Dover, New York, 1971, pp. 321-336.

Advanced Exposition

Abbott, J.C. The Chauvenet Papers: A Collection of Prize-Winning Expository Papers in Mathematics, 2 vols. Mathematical Association of America, Washington, D.C., 1978.

Aleksandrov, A.D., Kolmogorov, A.N., Lavrent′ev, M.A. Mathematics, Its Content, Methods, and Meaning, 3 vols. MIT Press, Cambridge, Massachusetts, 1969.

Behnke, H., et al. Fundamentals of Mathematics, 3 vols. MIT Press, Cambridge, Massachusetts, 1974.

Saaty, Thomas L. Lectures on Modern Mathematics, 3 vols. John Wiley, New York, 1963-1965.

Biography and Autobiography

Bell, Eric Temple. Men of Mathematics. Simon and Schuster, New York, 1937.

Box, Joan Fisher. R.A. Fisher, The Life of a Scientist. Wiley, New York, 1978.

Dauben, Joseph Warren. Georg Cantor: His Mathematics and Philosophy of the Infinite. Harvard University Press, Cambridge, Massachusetts, 1979.

Grattan-Guinness, Ivor. Joseph Fourier, 1768-1830. MIT Press, Cambridge, Massachusetts, 1972.

Halmos, Paul R. "Nicolas Bourbaki." Scientific American 196 (May 1957) 88-99, 174.

Halmos, Paul R. "The legend of John von Neumann." Amer. Math. Monthly 80 (1973) 382-394.

Hardy, G.H. Ramanujan. Chelsea Publishing, New York, 1968.

Hoffman, Banesh. Albert Einstein, Creator and Rebel. Viking Press, New York, 1972.

Infeld, Leopold. Whom the Gods Love. Whittlesey House, 1948; NCTM, Reston, Virginia, 1978.

Kovalevskaya, Sofya. Sofya Kovalevskaya: A Russian Childhood. Springer-Verlag, New York, 1978.

Mahoney, Michael S. The Mathematical Career of Pierre de Fermat. Princeton University Press, Princeton, New Jersey, 1973.

Meschkowski, Herbert. Ways of Thought of Great Mathematicians. Holden-Day, San Francisco, California, 1964.

Morgan, Bryan. Men and Discoveries in Mathematics. Transatlantic Arts, Inc., Levittown, New York, 1972.

Morse, Philip M. In at the Beginnings: A Physicist′s Life. MIT Press, Cambridge, Massachusetts, 1977.

Ore, Oystein. Niels Henrik Abel, Mathematician Extraordinary. University of Minnesota Press, Minneapolis, Minnsota, 1957; Chelsea Press, New York, 1974.

Osen, Lynn M. Women in Mathematics. MIT Press, Cambridge, Massachusetts, 1974.

Perl, Teri. Math Equals: Biographies of Women Mathematicians and Related Activities. Addison-Wesley, Reading, Massachusetts, 1978.

Reid, Constance. Courant in Göttingen and New York. Springer-Verlag, New York, 1976.

Reid, Constance. Hilbert. Springer-Verlag, New York, 1970.

Ulam, S.M. Adventures of a Mathematician. Charles Scribner′s Sons, New York, 1976.

Wiener, Norbert. Ex-Prodigy. Simon and Schuster, New York, 1953; MIT Press, Cambridge, Massachusetts, 1964.

Wiener, Norbert. I Am a Mathematician. Doubleday, New York, 1956; MIT Press, Cambridge, Massachusetts, 1964.

History

Al-Daffa, Ali Abdullah. The Muslim Contribution to Mathematics. Humanities Press, Atlantic Highlands, New Jersey, 1977.

Bell, Eric Temple. The Development of Mathematics. McGraw-Hill, New York, 1945.

Boyer, Carl B. A History of Mathematics. John Wiley, New York, 1968.

Chace, Arnold Buffum. The Rhind Mathematical Papyrus. NCTM, Reston, Virginia, 1979.

Goldstine, H. The Computer from Pascal to von Neumann. Princeton University Press, Princeton, New Jersey, 1972.

Kline, Morris. Mathematical Thought from Ancient to Modern Times. Oxford University Press, New York, 1972.

Kline, Morris. Mathematics in Western Culture. Oxford University Press, New York, 1953; 1964.

Kramer, Edna E. The Nature and Growth of Modern

Mathematics. Hawthorn, New York, 1970; Fawcett, New York, 1973.

Lambert, Joseph B., et al. "Maya arithmetic." Amer. Scientist 68 (May-June 1980) 249-255.

LeVeque, William J., et al. "History of mathematics." Encyclopaedia Britannica, 15th ed., 1974, Macropaedia, V. 11, pp. 639-670.

Menninger, Karl. Number Words and Number Symbols, A Cultural History of Numbers. MIT Press, Cambridge, Massachusetts, 1977.

Resnikoff, H.L. and Wells, R.O., Jr. Mathematics in Civilization: Geometry and Calculation as Keystones of Culture. Holt, Rinehart and Winston, New York, 1973.

Weyl, Hermann. "A half-century of mathematics." Amer. Math. Monthly 58 (1951) 523-553.

Wilder, Raymond L. "The origin and growth of mathematical concepts." Bull. Amer. Math. Soc. 59 (1963) 423-448.

Wilder, Raymond L. Evolution of Mathematical Concepts. Halsted Press, New York, 1974.

Wilder, Raymond L. "History in the mathematics curriculum: Its status, quality, and function." Amer. Math. Monthly 79 (1972) 479-495.

Zaslavsky, Claudia. Africa Counts: Number and Pattern in African Culture. Prindle, Weber and Schmidt, Boston, Massachusetts, 1973.

Mathematics and Science

Birkhoff, George D. "The mathematical nature of physical theories." Amer. Scientist 31 (1943) 281-310.

Browder, Felix E. "Is mathematics relevant? And if so, to what?" University of Chicago Magazine 67:3 (Spring 1975) 11-16; also appears as "The relevance of mathematics." Amer. Math. Monthly 83 (1976) 249-254.

Calder, Nigel. Einstein's Universe. Viking Press, New York, 1979.

Courant, Richard. "Mathematics in the modern world." Scientific American 211 (September 1964) 40-49, 269; also in M. Kline, Mathematics in the Modern World. W.H. Freeman, San Francisco, California, 1968, pp. 19-27, 394.

De Broglie, Louis. "The role of mathematics in the development of contemporary theoretial physics." in F. LeLionnais, Great Currents on Mathematical Thought, V. 2, Dover, New York, 1971, pp. 78-93.

Gardner, Martin. The Ambidextrous Universe: Mirror Asymmetry and Time-Reversed Worlds, Second Revised, Updated Edition. Scribner's, New York, 1979.

Penrose, Roger. "Einstein's vision and the mathematics of the natural world." The Sciences 19 (March 1979) 6-9.

Pólya, George. Mathematical Methods in Science. Mathematical Association of America, Washington, D.C., 1977.

Schwartz, Jacob T. "The pernicious influence of mathematics on science." in E. Nagel, P. Suppes and A. Tarski, Logic, Methodology and Philosophy of Science. Stanford University Press, Stanford, California, 1962, pp. 356-360.

Stone, Marshall H. "Mathematics and the future of science." Bull. Amer. Math. Soc. 63 (1957) 61-76.

Suppes, P. "A comparison of the meaning and uses of models in mathematics and the empirical sciences." Synthese 12 (1960) 287-301.

Wigner, Eugene P. "The unreasonable effectiveness of mathematics in the natural sciences." Comm. Pure Appl. Math. 13 (1960) 1-14; also in T.L. Saaty and F.J. Weyl. The Spirit and Uses of the Mathematical Sciences. McGraw-Hill, New York, 1969, pp. 123-140; in Studies in Mathematics, V. 16, SMSG, Stanford, California, 1967, pp. 31-44; and in E.P. Wigner, Symmetries and Reflections: Scientific Essays of Eugene P. Wigner, Indiana University Press, Bloomington, Indiana, 1967, pp. 222-237.

Zukav, Gary. The Dancing Wu Li Masters: An Overview of the New Physics. William Morrow, New York, 1979.

Mathematics and Society

Booss, Bernhelm and Niss, Mogens (eds.). Mathematics and the Real World. Birkhäuser, Boston, Massachusetts, 1979.

Fehr, Howard F. "Value and the study of mathematics." Scripta Math. 21 (1955) 49-53.

Morse, Marston. "Mathematics in our culture." in T.L. Saaty and F.J. Weyl, The Spirit and Uses of the Mathematical Sciences. McGraw-Hill, New York, 1969, pp. 105-120.

Whitehead, Alfred North. "Mathematics and liberal education." in A.N. Whitehead, Essays in Science and Philosophy. Philosophical Library, New York, 1947, pp. 175-188.

Whitehead, Alfred North. "Mathematics as an element in the history of thought." in J.R. Newman, The World of Mathematics, V. 1. Simon and Schuster, New York, 1956, pp. 402-416.

Wilder, Raymond L. "Trends and social implications of research." Bull. Amer. Math. Soc. 75 (1969) 891-906.

Philosophy and Logic

Baum, Robert J. Philosophy and Mathematics: From Plato to the Present. Freeman, Cooper and Company, San Francisco, California, 1974.

Benacerraf, Paul and Putnam, Hilary. Philosophy of Mathematics: Selected Readings. Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

Crossley, J.N., et al. What is Mathematical Logic? Oxford University Press, New York, 1972.

Hofstadter, Douglas R. Gödel, Escher, Bach: An Eternal Golden Braid. Basic Books, New York, 1979.

Lakatos, Imre. "Proofs and refutations." Brit. J. Phil. Science 14 (1963-64) 1-25, 120-139, 221-245, 296-342. Also available as: Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge University Press, New York, 1976.

Nagel, Ernest and Newman, James R. Gödel's Proof. New York University Press, New York, 1958.

Wilder, Raymond L. "The nature of mathematical proof." Amer. Math. Monthly 51 (1944) 309-323.

Wilder, Raymond L. Introduction to the Foundations of Mathematics, 2nd ed. John Wiley, New York, 1965.

Symmetry, Art, Aesthetics

Baker, Lillian F.; Schattschneider, Doris J. The Perspective Eye: Art and Math. Allentown Art Museum, Allentown, Pennsylvania, 1979.

Bezuszka, Stanley; Kenney, Margaret; and Silvey, Linda. Tessellations: The Geometry of Patterns. Creative Publications, California, 1977.

Birkhoff, George D. Aesthetic Measure. Harvard University Press, Cambridge, Massachusetts, 1933.

deFinetti, Bruno. Die Kunst des Sehens in der Mathematik. Birkhäuser, Basel, Switzerland, 1974.

Holden, Alan. Shapes, Space, and Symmetry. Columbia University Press, New York, 1971.

Huntley, H.E. The Divine Proportion: A Study in Mathematical Beauty. Dover, New York, 1970.

Lanczos, Cornelius. Space Through the Ages. Academic Press, New York, 1970.

Linn, Charles F. The Golden Mean: Mathematics and the Fine Arts. Doubleday and Company, New York, 1974.

Lockwood, E.H. and Macmillan, R.H. Geometric Symmetry. Cambridge University Press, New York, 1978.

Loeb, Arthur L. Space Structures, Their Harmony and Counterpoint. Addison-Wesley, Reading, Massachusetts, 1976.

Malina, Frank J., ed. Visual Art, Mathematics, and Computers: Selections from the Journal Leonardo. Pergamon, Elmsford, New York, 1979.

Mandelbrot, Benoit. Fractals: Form, Chance, and Dimension. W.H. Freeman, San Francisco, California, 1977.

Moineau, J.-C. Mathématique de l'esthétique. Dunod, Paris, France, 1969.

Ouchi, Hajime. Japanese Optical and Geometrical Art. Dover, New York, 1977.

Pearce, Peter. Structure in Nature as a Strategy for Design. MIT Press, Cambridge, Massachusetts, 1978.

Pearce, Peter and Pearce, Susan. Polyhedra Primer. D. Van Nostrand, New York, 1978.

Pedoe, Dan. Geometry and the Liberal Arts. St. Martin's Press, New York, 1978.

Pugh, Anthony. Polyhedra, A Visual Approach. University of California Press, California, 1976.

Robson, Ernest and Wimp, Jet, eds. Against Infinity: An Anthology of Contemporary Mathematical Poetry. Primary Press, Pennsylvania, 1979.

Rosen, Joe. Symmetry Discovered: Concepts and Applications in Nature and Science. Cambridge University Press, New York, 1975.

Senechal, Marjorie and Fleck, George. Patterns of Symmetry. University of Massachusetts Press, Amherst, Massachusetts, 1977.

Shubnikov, A.V. and Koptsik, V.A. Symmetry in Science and Art. Plenum Press, New York, 1974.

Stevens, Peter S. Patterns in Nature. Little, Brown and Company, Boston, Massachusetts, 1974.

Wechsler, Judith, ed. On Aesthetics in Science. MIT Press, Massachusetts, 1978.

Weyl, Hermann. Symmetry. Princeton University Press, Princeton, New Jersey, 1952; excerpted in J.R. Newman. The World of Mathematics, V. 1. Simon and Schuster, New York, 1956, pp. 671-724.

Computing

Feldman, Jerome A. "Programming languages." Scientific American 241 (December 1979) 94-116.

Hamming, Richard W. "Intellectual implications o the computer revolution." Amer. Math. Monthly 70 (1963) 4-11; also in T.L. Saaty and F.J. Weyl. The Spirit and Uses of the Mathematical Sciences, McGraw-Hill, 1969, pp. 188-199; in Studies in Mathematics, V. 16, School Mathematics Study Group, Stanford, California, 1967, pp. 45-52; and in Z.W. Pylyshyn. Perspectives on the Computer Revolution. Prentice-Hall, Englewood Cliffs, New Jersey, 1970, pp. 370-377.

McCorduck, Pamela. Machines Who Think: A Persona Inquiry into the History of Prospects of Artificial Intelligence. Freeman, San Francisco, California, 1979.

Weizenbaum, Joseph. Computer Power and Human Reason: From Judgment to Calculation. W.H. Freeman, San Francisco, California, 1976.

Pedagogy

Dieudonné, Jean A. "Should we teach modern mathematics?" Amer. Scientist 61 (1973) 16-19.

Engel, Arthur. "The relevance of modern fields o applied mathematics for mathematical education." Educ. Studies Math. 2 (1969-70) 257-269.

Henrici, Peter. "Reflections of a teacher of applied mathematics." Quarterly of Applied Math. 30 (1972) 31-39.

Hilton, Peter J. "The survival of education." Educ. Tech. 13:11 (November 1973) 12-16.

Kline, Morris. Why the Professor Can't Teach: Mathematics and the Dilemma of University Education. St. Martin's Press, New York, 1977.

Kline, Morris. "Logic versus pedagogy." Amer. Math. Monthly 77 (1970) 264-282.

Klamkin, Murray S. "The teaching of mathematics so as to be useful." Educ. Studies Math. 1 (1968-69) 126-160.

Kemeny, John G. "Teaching the new mathematics." Atlantic Monthly 210 (October 1962) 90-91+; also in J.G. Kemeny. Random Essays on Mathematics, Education and Computers. Prentice-Hall, Englewooc Cliffs, New Jersey, 1964, pp. 27-34.

Lazarus, Mitchell. "Mathophobia: Some personal speculations." Nat. Elem. Principal 53:2 (Jan.-Feb. 1974) 16-22.

Lighthill, M.J. "The art of teaching the art of applying mathematics." Math. Gazette 55 (1971) 249-270.

Ordman, Edward T. "One and one is nothing: Liberating mathematics." Soundings 56 (1973) 164-181.

Pollak, Henry O. "How can we teach applications of math?" Educ. Studies Math. 2 (1969-70) 393-404.

Pollak, Henry O. "On some of the problems of teaching applications of mathematics." Educ. Studies Math. 1 (1968-69) 24-30.

Pólya, George. How to Solve It. Princeton University Press, Princeton, New Jersey, 1945; excerpted in J.R. Newman. The World of

Mathematics, V. 3. Simon and Schuster, New York, 1956, pp. 1980-1992.

Pólya, George. Mathematical Discovery, 2 vols. John Wiley, New York, 1962 and 1965.

Pólya, George. Mathematics and Plausible Reasoning, Vols. I and II. Princeton University Press, Princeton, New Jersey. Vol. I, 1954; Vol. II, rev. ed., 1969.

Puzzles & Recreations

Ball, W.W. Rouse and Coxeter, H.S. MacDonald. Mathematical Recreations and Essays, Twelfth Edition. University of Toronto Press, Toronto, Canada, 1974.

Dudeney, Henry Ernest. The Canterbury Puzzles and Other Curious Problems, Fourth Edition. Dover, New York, 1958.

Duffin, R.J. Puzzles, Games, and Paradoxes. Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1979.

Fixx, James E. Solve It! A Perplexing Profussion of Puzzles. Doubleday, New York, 1978.

Fujimura, Kobon. The Tokyo Puzzles. Charles Scribner's Sons, New York, 1978.

Gardner, Martin. Martin Gardner's Sixth Book of Mathematical Games from Scientific American. W.H. Freeman, San Francisco, California, 1971.

Gardner, Martin. Mathematical Carnival. Alfred A. Knopf, Inc., New York, 1975.

Gardner, Martin. Mathematical Circus. Alfred A. Knopf, Inc., New York, 1979.

Gardner, Martin. Mathematical Magic Show. Alfred A. Knopf, New York, 1977.

Gardner, Martin. Mathematics, Magic and Mystery. Dover, New York, 1956.

Gardner, Martin. New Mathematical Diversions from Scientific American. Simon and Schuster, New York, 1966.

Gardner, Martin. The Numerology of Dr. Matrix. Simon and Schuster, New York, 1967.

Gardner, Martin. The Scientific American Book of Mathematical Puzzles and Diversions. Simon and Schuster, New York, 1959.

Gardner, Martin. The Second Scientific American Book of Mathematical Puzzles and Diversions. Simon and Schuster, New York, 1961.

Gardner, Martin. The Unexpected Hanging, and Other Mathematical Diversions. Simon and Schuster, New York, 1969.

Hunter, J.A.H. Mathematical Brain-Teasers. Dover, New York, 1976.

Hunter, J.A.H. and Madachy, Joseph S. Mathematical Diversions. Dover, New York, 1975.

Kraitchik, Maurice. Mathematical Recreations, 2nd ed. Dover, New York, 1953.

Mott-Smith, Geoffrey. Mathematical Puzzles for Beginners and Enthusiasts, Second Revised Edition. Dover, New York, 1954.

Ogilvy, C. Stanley. Tomorrow's Math: Unsolved Problems for the Amateur, Second Edition. Oxford University Press, New York, 1972.

Schwartz, Benjamin L., ed. Mathematical Solitaires & Games. Baywood Publishers, Farmingdale, New York, 1980.

Singmaster, David. Notes on Rubik's Magic Cube, Fifth Edition, Preliminary Version. Polytechnic of the South Bank, London, 1980.

Smullyan, Raymond. The Chess Mysteries of Sherlock Holmes. Alfred A. Knopf, New York, 1979.

Smullyan, Raymond M. What is the Name of This Book? Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

Tietze, Heinrich. Famous Problems of Mathematics. Graylock Press, Baltimore, Maryland, 1965.

Reference

Gaffney, Matthew P. and Steen, Lynn Arthur. Annotated Bibliography of Expository Writing in the Mathematical Sciences. Mathematical Association of America, Washington, D.C., 1976.

Hoyrup, Else. Books About Mathematics: History, Philosophy, Education, Models, System Theory, and Works of Reference, etc.: A Bibliography. Roskilde University Center, Denmark, 1979.

Hoyrup, Else. Women and Mathematics, Science and Engineering. Roskilde University Center, Denmark, 1978.

A Basic Library List for Four-Year Colleges, Second Edition. Mathematical Association of America, Washington, D.C., 1976.

A Basic Library List for Two-Year Colleges, Second Edition. Mathematical Association of America, Washington, D.C., 1980.

May, Kenneth O. Bibliography and Research Manual of the History of Mathematics. University of Toronto Press, Toronto, Canada, 1973.

May, Kenneth O. Index of the American Mathematical Monthly, V. 1-80. (1894-1973). Mathematical Association of America, Washington, D.C., 1977.

Schaaf, William L. A Bibliography of Recreational Mathematics, V. 1-4. National Council of Teachers of Mathematics, Reston, Virginia, 1954, 1970, 1973, 1978..

Schaaf, William L. Mathematics and Science: An Adventure in Postage Stamps. National Council of Teachers of Mathematics, Reston, Virginia, 1978.

Schaefer, Barbara Kirsch. Using the Mathematical Literature, A Practical Guide. Dekker, New York, 1979.

Schneider, David. Annotated Bibliography of Films and Videotapes for College Mathematics. Mathematical Association of America, Washington, D.C., 1980.

Seebach, J. Arthur and Steen, Lynn Arthur. Mathematics Magazine: 50 Year Index (1926-1977). Mathematical Association of America, Washington, D. C., 1979.

Singmaster, David. List of 16mm Films on Mathematical Subjects. Open University, England.

Fiction, Fables and Anecdotes

Abbott, Edwin A. Flatland--A Romance of Many Dimensions. Little, Brown, Boston, Massachusetts, 1928; Dover, New York, 1952.

Eves, Howard W. In Mathematical Circles. Prindle, Weber and Schmidt, Boston, Massachusetts, 1969.

Eves, Howard W. Mathematical Circles Revisited. Prindle, Weber and Schmidt, Boston, Massachusetts,

1971.

Eves, Howard W. Mathematical Circles Squared. Prindle, Weber and Schmidt, Boston, Massachusetts, 1971.

Eves, Howard W. Mathematical Circles Adieu. Prindle, Weber and Schmidt, Boston, Massachusetts, 1977.

Fadiman, Clifton. Fantasia Mathematica. Simon and Schuster, New York, 1961.

Fadiman, Clifton. The Mathematical Magpie. Simon and Schuster, New York, 1962.

Moritz, Robert Edouard. On Mathematics: A Collection of Witty, Profound, Amusing Passages About Mathematics and Mathematicians. Dover, New York, 1942.

Probability and Statistics

David, F.N. Games, Gods and Gambling. Hafner Press, New York, 1962.

Huff, Darrel and Geis, Irving. How to Lie with Statistics. W.W. Norton, New York, 1954.

Kimble, Gregory A. How to Use (and Misuse) Statistics. Prentice-Hall, New Jersey, 1978.

Levinson, Horace C. Chance, Luck, and Statistics, Second Edition. Dover, New York, 1963.

Moore, David S. Statistics: Concepts and Controversies. Freeman, San Francisco, California, 1979.

Tanur, Judith M., ed. Statistics: A Guide to the Unknown, Second Edition. Holden-Day, California, 1978.

Williams, Bill. A Sampler on Sampling. Wiley, New York, 1978.

Topology and Geometry

Barr, Stephen. Experiments in Topology. Thomas Y. Crowell, New York, 1972.

Engel, Kenneth. "Shadows of the 4th dimension." Science 80 (July/August 1980) 68-73.

Flegg, H. Graham. From Geometry to Topology. English University Press (Crane Rusak, New York, distributor), 1974.

Gray, Jeremy. Ideas of Space: Euclidean, Non-Euclidean, and Relativistic. Clarendon Press, New York, 1979.

Griffiths, H.B. Surfaces. Cambridge University Press, New York, 1976.

Pedoe, D. Circles, A Mathematical View. Dover, New York, 1979.

Wenninger, Magnus J. Polyhedron Models. Cambridge University Press, New York, 1971.

Wenninger, Magnus J. Spherical Models. Cambridge University Press, New York, 1979.

Miscellaneous Books

Asimov, Isaac. Asimov on Numbers. Doubleday and Company, New York, 1977.

Brams, Steven J. Spatial Models of Election Competition. EDC/UMAP, Newton, Massachusetts, 1979.

Brams, Steven J. Biblical Games: A Strategic Analysis of Stories in the Old Testament. MIT Press, Cambridge, Massachusetts, 1980.

Beck, Anatole, Bleicher, Michael N. and Crowe, Donald W. Excursions Into Mathematics. Worth Publishers, New York, 1969.

De Morgan, Augustus. A Budget of Paradoxes. Open Court, LaSalle, Illinois, 1872; 1915; excerpted in J.R. Newman. The World of Mathematics, V. 4. Simon and Schuster, New York, 1956, pp. 2369-2382.

Honsberger, Ross. Mathematical Gems from Elementary Combinatorics, Number Theory, and Geometry. Mathematical Association of America, Washington, D.C., 1973.

Honsberger, Ross. Ingenuity in Mathematics. Mathematical Association of America, Washington, D.C., 1975.

Honsberger, Ross. Mathematical Gems II. Mathematical Association of America, Washington, D.C., 1976.

Honsberger, Ross. Mathematical Morsels. Mathematical Association of America, Washington, D.C., 1979.

Honsberger, Ross, ed. Mathematical Plums. Mathematical Association of America, Washington, D.C., 1979.

Kac, Mark and Ulam, Stanislaw M. Mathematics and Logic: Retrospect and Prospects. Frederick A. Praeger, New York, 1969.

Knuth, D.E. Surreal Numbers. Addison-Wesley, Reading, Massachusetts, 1974.

Kogelman, Stanley; Warren, Joseph. Mind over Math. Dial Press, New York, 1978. See also: Hilton, Peter and Pedersen, Jean. "Review of 'Overcoming Math Anxiety' and 'Mind over Math'." Amer. Math. Monthly 87 (1980) 143-148.

Lieber, Lillian R. Human Values and Science, Art and Mathematics. W.W. Norton and Company, New York, 1961.

Lieber, Lillian R. Mits, Wits, and Logic. W.W. Norton, New York, 1947; 1954; 1960.

Lieber, Lillian R. Take a Number. Ronald Press, New York, 1946.

Lieber, Lillian R. The Education of T.C. Mits. W.W. Norton, New York, 1942.

Melzak, Z.A. Companion to Concrete Mathematics, 2 vols. Wiley, New York, 1973, 1976.

Paulos, John Allen. Mathematics and Humor. University of Chicago Press, Chicago, Illinois, 1980.

Roberts, Fred S. Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems. Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

Rubinstein, Moshe F. Patterns of Problem Solving. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

Tobias, Sheila. Overcoming Math Anxiety. Norton, New York, 1978. See also: Hilton, Peter and Pedersen, Jean. "Review of 'Overcoming Math Anxiety' and 'Mind over Math'." Amer. Math. Monthly 87 (1980) 143-148.

Waisman, Friedrich. Introduction to Mathematical Thinking: The Formation of Concepts in Modern Mathematics. Harper & Brothers, New York, 1959.

Woodcock, Alexander and Davis, Monte. Catastrophe Theory. E.P. Dutton and Company, New York, 1978.

8. C. J. Everett, Iteration of the number-theoretic function $f(2n) = n, f(2n+1) = 3n+2$, Adv. in Math., 25 (1977) 42–45; MR **56** #15552; Zbl. 352.10001.

9. G. Frobenius, Über die Markoffschen Zahlen, S.-B. Preuss. Akad. Wiss. Berlin (1913) 458–487.

10. Martin Gardner, Mathematical Games, A miscellany of transcendental problems, simple to state but not at all easy to solve, Scientific Amer., 226 # 6(Jun 1972) 114–118, esp. p. 115.

11. Martin Gardner, Mathematical Games, Patterns in primes are a clue to the strong law of small numbers, Scientific Amer., 243 #6 (Dec 1980) 18–28.

12. Richard K. Guy, Unsolved Problems in Number Theory, Springer, New York, 1981, Problems D12, E16, E17.

13. E. Heppner, Eine Bemerkung zum Hasse-Syracuse-Algorithmus, Arch. Math. (Basel), 31 (1977/79) 317–320; MR **80**d: 10007; Zbl. 377.10027.

14. David C. Kay, Pi Mu Epsilon J., 5 (1972) 338.

15. A Markoff, Sur les formes quadratiques binaires indéfinies, Math. Ann., 15 (1879) 381–409.

16. H. Möller, Über Hasses Verallgemeinerung der Syracuse-Algorithmus (Kakutani's Problem), Acta. Arith., 34 (1978) 219–226; MR **57** #16246; Zbl. 329.10008.

17. R. Remak, Über indefinite binäre quadratische Minimalformen, Math. Ann., 92 (1924) 155–182.

18. R. Remak, Über die geometrische Darstellung der indefinitiven binären quadratischen Minimalformen, Jber. Deutsch Math.-Verein, 33 (1925) 228–245.

19. Gerhard Rosenberger, The uniqueness of the Markoff numbers, Math. Comp., 30 (1976) 361–365; but see MR **53** #280.

20. Ray P. Steiner, A theorem on the Syracuse problem, Congressus Numerantium XX, Proc. 7th Conf. Numerical Math. Comput. Manitoba, 1977, 553–559; MR **80**g:10003.

21. Riho Terras, A stopping time problem on the positive integers, Acta Arith., 30 (1976) 241–252; MR 58 #27879 (and see 35 (1979) 100–102; MR **80**h:10066).

22. L. Ja. Vulah, The diophantine equation $p^2 + 2q^2 + 3r^2 = 6pqr$ and the Markoff spectrum (Russian), Trudy Moskov. Inst. Radiotehn. Elektron. i Avtomat. Vyp. 67 Mat. (1973) 105–112, 152; MR **58** #21957.

23. Don B. Zagier, Distribution of Markov numbers, Abstract 796–A37, Notices Amer. Math. Soc., 26 (1979) A-543.

24. David A. Klarner, An algorithm to determine when certain sets have 0-density, J. Algorithms, 2 (1981) 31–43; Zbl. 464.10046.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, NM 55057.

WELL-DISTRIBUTED MEASURABLE SETS

WALTER RUDIN

Department of Mathematics, University of Wisconsin, Madison, WI 53706

THEOREM. *There is a measurable set $A \subset I = [0, 1]$ such that*

$$0 < m(A \cap V) < m(V)$$

for every nonempty open set $V \subset I$.

Proof. Let CTDP mean: Compact Totally Disconnected subset of I, having Positive measure.

Let $\{I_n\}$ be an enumeration of all segments in I whose endpoints are rational. Construct sequences $\{A_n\}$, $\{B_n\}$ of CTDP's as follows:

Start with disjoint CTDP's A_1 and B_1 in I_1.

Once $A_1, B_1, \ldots, A_{n-1}, B_{n-1}$ are chosen, their union C_n is CTD, hence $I_n \setminus C_n$ contains a

nonempty segment J, and J contains a pair A_n, B_n of disjoint CTDP's. Continue in this way, and put

$$A = \bigcup_{n=1}^{\infty} A_n.$$

If $V \subset I$ is open and nonempty, then $I_n \subset V$ for some n, hence $A_n \subset V$ and $B_n \subset V$. Thus

$$0 < m(A_n) \leq m(A \cap V) < m(A \cap V) + m(B_n) \leq m(V);$$

the last inequality holds because A and B_n are disjoint. Done.

The point of publishing this is to show that the highly computational construction of such a set in [**1**] is much more complicated than necessary.

Reference

1. A. Simoson, An Archimedean paradox, this MONTHLY, 89 (1982) 114–125.

ANY QUESTIONS?

DESMOND MACHALE
Department of Mathematics, University College, Cork, Ireland

Recently, I attended a mathematical lecture given by a guest speaker where absolutely nobody, except possibly the speaker, had the remotest idea what was going on. Normally, one can absorb at least some of the preliminary definitions and follow, say, the first blackboard full of development of the theory, but on this occasion everyone was completely lost after the first definition. After the speaker had finished over an hour later to an enthusiastic round of applause, the chairman asked for questions, and, of course, there was a deathly and highly embarrassing silence. Then and there I resolved to put together a collection of universal questions for use in such situations. Such questions must sound sensible, but they are designed to cover up the total ignorance of the questioner rather than to elicit information from the speaker. The following is the list I came up with.

1. Can you produce a series of counterexamples to show that if any of the conditions of the main theorem are dropped or weakened, then the theorem no longer holds?
   [The speaker can almost always do so—if not you may have presented him with a stronger theorem!]
2. What inadequacies of the classical treatment of this subject are now becoming obvious?
3. Can your results be unified and generalized by expressing them in the language of Category Theory?
   [The answer to this question is always NO!]
4. Isn't there a suggestion of Theorem 3 in an early paper of Gauss?
   [The answer to this question is almost always YES!]
5. Isn't the constant 4.15 in Theorem 2 suspiciously close to $4\pi/3$?
   [This question can clearly be generalized for any constant k—"Isn't k suspiciously close to $(p/q)\pi$ (for suitable integers p and q)?"]
6. I'm not sure I understand the proof of Lemma 3—could you outline it for us again?
   [Lemma 3 should be just a little nontrivial, yet not more than one third of a blackboard in length.]
7. Are you familiar with a joint paper of Besovik and Bombialdi which might explain why the converse of Theorem 5 is false without further assumptions?
   [This is a dangerous question to ask unless you like living dangerously. The answer is always "NO" unless the speaker is playing the same game as you are, because Besovik and

nonempty segment J, and J contains a pair A_n, B_n of disjoint CTDP's. Continue in this way, and put

$$A = \bigcup_{n=1}^{\infty} A_n.$$

If $V \subset I$ is open and nonempty, then $I_n \subset V$ for some n, hence $A_n \subset V$ and $B_n \subset V$. Thus

$$0 < m(A_n) \le m(A \cap V) < m(A \cap V) + m(B_n) \le m(V);$$

the last inequality holds because A and B_n are disjoint. Done.

The point of publishing this is to show that the highly computational construction of such a set in [**1**] is much more complicated than necessary.

Reference

1. A. Simoson, An Archimedean paradox, this MONTHLY, 89 (1982) 114–125.

ANY QUESTIONS?

DESMOND MACHALE
Department of Mathematics, University College, Cork, Ireland

Recently, I attended a mathematical lecture given by a guest speaker where absolutely nobody, except possibly the speaker, had the remotest idea what was going on. Normally, one can absorb at least some of the preliminary definitions and follow, say, the first blackboard full of development of the theory, but on this occasion everyone was completely lost after the first definition. After the speaker had finished over an hour later to an enthusiastic round of applause, the chairman asked for questions, and, of course, there was a deathly and highly embarrassing silence. Then and there I resolved to put together a collection of universal questions for use in such situations. Such questions must sound sensible, but they are designed to cover up the total ignorance of the questioner rather than to elicit information from the speaker. The following is the list I came up with.

1. Can you produce a series of counterexamples to show that if any of the conditions of the main theorem are dropped or weakened, then the theorem no longer holds?
   [The speaker can almost always do so—if not you may have presented him with a stronger theorem!]
2. What inadequacies of the classical treatment of this subject are now becoming obvious?
3. Can your results be unified and generalized by expressing them in the language of Category Theory?
   [The answer to this question is always NO!]
4. Isn't there a suggestion of Theorem 3 in an early paper of Gauss?
   [The answer to this question is almost always YES!]
5. Isn't the constant 4.15 in Theorem 2 suspiciously close to $4\pi/3$?
   [This question can clearly be generalized for any constant k—"Isn't k suspiciously close to $(p/q)\pi$ (for suitable integers p and q)?"]
6. I'm not sure I understand the proof of Lemma 3—could you outline it for us again?
   [Lemma 3 should be just a little nontrivial, yet not more than one third of a blackboard in length.]
7. Are you familiar with a joint paper of Besovik and Bombialdi which might explain why the converse of Theorem 5 is false without further assumptions?
   [This is a dangerous question to ask unless you like living dangerously. The answer is always "NO" unless the speaker is playing the same game as you are, because Besovik and

Bombialdi do not exist, and even if by some unfortunate chance they do exist, it is very unlikely that they have written a joint paper. If the speaker calls your bluff and asks for details and a reference, tell him the paper is available only in Albanian with Portuguese summaries. Promise to mail him a copy but forget to do.]

8. Why not get a graduate student to perform the horrendous calculations mentioned in Theorem 1 in the case $n = 4$?
   [The answer is always "I've a student doing just that at the moment."]
9. Could you draw us a simple diagram to show what the situation looks like for $n = 2$?
   [Be careful that he hasn't already done so.]
10. What textbook would you recommend for someone who wishes to get students interested in this area?
    [The speaker has almost invariably written such a textbook himself and will be delighted you asked this question. If he hasn't, then you can ask the next question.]
11. When can we expect your definitive textbook on this subject?
12. Why do you think there was such a flurry of activity in this area around the turn of the century and then nothing until your paper of 1979?
    [The true answer is that people in the period in between had more sense.]

In general, a good ploy is to stop halfway through a totally meaningless question you are asking and pretend you have suddenly seen the answer yourself. However, never, *never* ask

13. What are the applications of these results?

The speaker is probably embarrassed enough already!

AN ALGORITHM FOR THE MINIMAL POLYNOMIAL OF A MATRIX

BERNARD R. GELBAUM
Department of Mathematics, State University of New York at Buffalo, NY 14214

Using nothing beyond the Gram-Schmidt orthonormalization process, we present a direct method for finding the minimal polynomial of a matrix. We will use the following notation: $A = (a_{ij})$ is an $n \times n$ complex matrix;

$$m_A(\lambda) = \lambda^m + a_1\lambda^{m-1} + \cdots + a_m,$$

the unique monic polynomial of least degree such that $m_A(A)\colon A^m + \sum_{i=1}^{m} a_i A^{m-i} = 0$ is the minimal polynomial for A. M_n will denote the vector space of $n \times n$ complex matrices. M_n has dimension n^2 and hence is isomorphic to C^{n^2}, the vector space of n^2-tuples of complex numbers. We will use the standard isomorphism $\phi : M_n \to C^{n^2}$ which spreads the rows of a matrix out in a long line:

$$\phi((a_{ij})) = (a_{11}, \cdots, a_{1n}, a_{21}, \cdots, a_{2n}, \cdots, a_{n1}, \cdots, a_{nn}).$$

We will also use the standard inner product for C^{n^2}: $\langle u, v\rangle = \sum_{i=1}^{n^2} u_i \bar{v}_i$ where $u = (u_i)$, $v = (v_i)$ and $\bar{v}_i$ denotes complex conjugation.

Now apply the Gram-Schmidt orthonormalization to the powers of a matrix A, considered as vectors in C^{n^2}, to obtain the coefficients involved in a minimal linear combination of these powers of A; i.e., the coefficients of the minimal polynomial. Let A^0 be the identity matrix I and recall that $\phi(A^p)$ is just A^p considered as a spread-out vector in C^{n^2}. Let O_p be the pth orthonormalized vector produced by the Gram-Schmidt process. Then $O_0 = \phi(I)/\sqrt{n}$ and O_{p+1} is found inductively by computing

$$X_{p+1} = \phi(A^{p+1}) - \sum_{k=0}^{p} < \phi(A^{p+1}), O_k > O_k$$

and then taking $O_{p+1} = X_{p+1}/\|X_{p+1}\|$. Since we are orthonormalizing in a vector space of finite dimension n^2, we eventually have $X_{m+1} = 0$ for some $m \leqslant n^2$. The Gram-Schmidt process stops here. Gather up the various coefficients computed in the orthonormalization process and express each O_k as a linear combination of $\phi(A^k), \cdots, \phi(A), \phi(I)$ and thereby obtain a relation

$$\phi(A^m) + \sum_{k=1}^{m} a_k \phi(A^{m-k}) = O.$$

Since the process stops at the first power of A that is linearly dependent on the preceding (lower) powers of A, it is clear that

$$m_A(\lambda) = \lambda^m + \sum_{k=1}^{m} a_k \lambda^{m-k}.$$

We close with a few brief observations. Although it is not hard to write a computer program for carrying out the above procedure, it is apparently not easy to compare the efficiency of this approach to that of using the formula $m_A(\lambda) = \det(A - \lambda I)/D_{n-1}(\lambda)$ where $D_{n-1}(\lambda) = \text{GCD}$ of all the $(n-1)$ st order minors of $\det(A - \lambda I)$. Finally if v is any vector in C^n the Gram-Schmidt process applied as above to the vectors $v, Av, A^2v, \ldots$ in C^n leads to the so-called *order* $m_{A,v}(\lambda)$ of v with respect to A, i.e., to the unique monic polynomial

$$\lambda^\mu + \sum_{k=1}^{\mu} b_k \lambda^{\mu-k}$$

of least degree μ such that $m_{A,v}(A)v = 0$. Clearly $\mu < n$ and then by the Cayley-Hamilton theorem $m \leqslant n$.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

MATHEMATICS APPRECIATION COURSES: THE REPORT OF A CUPM PANEL*

In 1977 the Committee on the Undergraduate Program in Mathematics (CUPM) established a panel to consider the content of those college and university courses that treat mathematics appreciation for students in the arts and humanities. Such courses are taken by a large number of students, frequently as their last formal contact with mathematics. Yet in most institutions they are given very low priority; they are frequently taught perfunctorily, without a clear set of objectives, by faculty who lack appropriate interest or credentials. Since these courses may play a major role in molding nonscientists' opinions of mathematics and its role in society, CUPM decided that it should call attention to the importance of these courses and offer some suggestions on how they may be organized and taught effectively.

This is the report, approved by CUPM, of the CUPM Panel on Mathematics Appreciation Courses. While the panel has many guidelines and recommendations to offer, it does not feel that a particular selection of topics or teaching strategy should be universally adopted for mathematics appreciation courses. A main goal of such courses is to get students to appreciate the significant role that mathematics plays in society, both past and present. All material presented in such courses should be well motivated and related to the role of mathematics in culture and technology.

*Reprints of this article and the related bibliography contained in the Center Section may be purchased for $1.00 from MAA Publications Department, 1529 Eighteenth Street, NW, Washington, D. C. 20036.

Acknowledgements

Many people, including nonmathematicians, responded to the Panel's request for information and advice. An open meeting of the panel at the MAA meeting in Biloxi, Mississippi, in January, 1979, was attended by more than sixty people who had much to contribute. The panel particularly wishes to thank Professors Henry Alder, Dorothy Bernstein, Donald Bushaw, William Lucas, David Penney, William M. Priestley, David Roselle, and Alice Schafer for their helpful comments.

Bettye Anne Case
John Conway
Richard J. Duffin
Jerome A. Goldstein, Chairman
Elaine Koppelman
Kenneth Rebman
Lynn Arthur Steen
James Vineyard

I. Mathematics Appreciation: A Philosophy. The inclusion of a mathematics appreciation course in the undergraduate curriculum is common in the nation's colleges and universities. This trend is a direct result of an underlying belief, held by most mathematicians, that every well-educated person should be mathematically literate. Whether or not a mathematics course is required at a particular institution often depends, among other things, upon the extent to which this belief is shared by the general faculty. But the ultimate success of an "appreciation" course in mathematics should not depend upon mandatory enrollment. Rather, the value and importance of such a course should be directly attributable to the care and understanding with which it is conceived and taught.

If as mathematicians we accept the notion that an educated person should know something about mathematics, then we must also accept the responsibility for conscientiously providing appropriate training. Students in the mathematical, physical, life, and some social sciences, and usually those in business, study mathematics as an inherent part of the undergraduate curriculum. It is not to these students, but rather to majors in the arts, in the humanities, and in certain social sciences that we must direct the mathematics appreciation course. At the outset we must take into account the background and interests of the prospective students. In many cases they have chosen their majors precisely because of a weak or unpleasant mathematical background; a college course that reinforces this negative experience with mathematics certainly cannot be called an appreciation course.

At most institutions the great majority of students in a mathematics appreciation course will have studied less than four years of high school mathematics; moreover, many of these students will have had poor experience in mathematics, or will have had very weak courses. However, high school mathematics study is predominantly concerned with developing skills, and while such skills are of unquestioned importance, they are not necessarily prerequisite to (nor should teaching them be a part of) a mathematics appreciation course.

Among all fundamental academic disciplines, mathematics is perhaps unique in the degree to which it is not understood (or is misunderstood) by students and even faculty from other areas of study. By taking an introductory course in chemistry, history, or psychology, a student is expected to gain an understanding of the general techniques, accomplishments, and goals of the discipline, and will learn to appreciate the work of the contemporary professional practitioner of the subject, sometimes even to the extent of reading the current journals. But an undergraduate major in mathematics is unlikely to have comparable insight into mathematics. Thus the challenge of a mathematics appreciation course is enormous.

The ultimate goal of such courses is defined by our umbrella title—to instill in the student an appreciation of mathematics. For this to occur, students must come to understand the historical and contemporary role of mathematics, and to place the discipline properly in the context of other human intellectual achievements.

From the beginning of recorded history, mathematics has proved to be an indispensable aid to the empirical sciences; the great successes (and failures) of mathematical reasoning in the furtherance of human knowledge are tales begging to be told. Even the direct impact of mathematics on developments in virtually all disciplines is often not realized by the mathematical layman.

But of course, to mathematicians, the subject is more than a tool of applied science, more than a universal language useful for communication and research in other disciplines. Mathematicians see mathematics as an intellectually exciting discipline, one that holds great aesthetic appeal for its practitioners. This idea of mathematics as art is often difficult for nonmathematicians to appreciate, yet is fundamental to understanding the development and role of the subject.

Finally, to appreciate mathematics fully, one must recognize it as a vital, on-going discipline, one that is practiced by a worldwide community of dedicated, sometimes passionate, and frequently brilliant scholars. It is a surprise to many that mathematics is a living, changing, developing subject. A true appreciation of mathematics requires some knowledge of contemporary developments.

The entire mathematical community should be concerned with what view educated, informed people have of mathematics. Thus, courses in mathematics appreciation, while presumably benefiting primarily the students, may also have a long-term positive effect on the discipline itself. Obvious benefits will accrue if leaders in education, industry, business, and government have a better understanding of the nature, role, and importance of contemporary mathematics.

It is a sad commentary on the attitudes of mathematicians that courses in mathematics appreciation frequently command pejorative (albeit informal) labels such as "Math for Poets." Even the supposedly neutral title of "Math for Liberal Arts Students" may convey the connotation of condescension. We must recall that liberal arts education, for a large percentage of the college educated population, is a rigorous, disciplined encounter with the best elements of man's history and culture. The major clientele of the mathematics appreciation courses are liberal arts students, and it is from their ranks that many of society's leaders will emerge.

The panel believes that it is better to describe courses of this type in terms of their objectives rather than their audience. Since the term "mathematics appreciation" brings to mind similar courses in other special fields (e.g., "music appreciation") that generally carry positive connotations with regard to their role in general undergraduate education, and since it conveys concisely what such courses intend to accomplish, standing as a brief reminder of this intention to both teachers and students, the majority of the Panel prefers this title.

II. Things to Stress in a Mathematics Appreciation Course.

1. *The relationship between mathematics and our cultural heritage*. Students enrolled in mathematics appreciation courses are generally more interested in, as well as more knowledgeable about, the arts and humanities than the sciences; it is natural, therefore, to capitalize on these strengths by appropriate illustrations of the relations between mathematics and music, art, literature, history, and society.

2. *The role of mathematics in history and the role of history in mathematics*. Although the influence of mathematics is often remote, mathematical discoveries have shaped our world in fundamental ways, altering the course of history as well as the way we live and work. Examples of these influences abound, and should form a major part of any mathematics appreciation course. Historical developments and the evolution of mathematical concepts should be properly emphasized.

3. *The nature of contemporary mathematics*. The mathematics known by most humanities students is ancient mathematics—the geometry of ancient Greece, and the algebra of the early renaissance; not surprisingly, such students have the impression that mathematics is dead. Showing them that it is in fact a vigorous, growing discipline with considerable influence in contemporary society is an important aspect of any course in mathematics appreciation.

4. *The recent emergence of several mathematical sciences*. While the mathematics appreciation course should not be devoted solely to one "modern" area such as statistics, computer science, or operations research, it surely provides an opportunity to use these fields as illustrations of the panoply of contemporary mathematical science.

5. *The necessity of doing mathematics to learn mathematics.* While some parts of the mathematics appreciation course can and should be about mathematics, it is essential that some parts actually engage the students in doing mathematics. Only in this way can they gain a realistic sense of the process and nature of mathematics. Of course it is vitally important that the instructor have appropriate respect for the students' interests and abilities, and that exercises be selected so as to maintain rather than destroy their enthusiasm.

6. *The role of mathematics as a tool for problem solving.* As the language of science and industry, mathematical models are the tool *par excellence* for solving problems. Students in mathematics appreciation courses should be exposed to contemporary mathematical modelling, to gain some appreciation both of its power and its limitations.

7. *The verbalization and reasoning necessary to understand symbolism.* While symbols provide the mathematician and scientist with great power, they obscure the meaning of mathematics from the uninitiated. A great service the teacher of a mathematics appreciation course can provide is to enable students to overcome their fear of symbols, to learn to think through arguments apart from the traditional symbols in which they are expressed.

8. *The existence of a large body of interesting writing about mathematics.* Students in mathematics appreciation courses generally feel comfortable with assignments such as term papers, book reports, and library research because they have become accustomed to these in their humanities courses. There is much good mathematics that can be learned in this way, and assignments can be arranged that utilize these familiar learning tools.

III. Things to Avoid in a Mathematics Appreciation Course.

1. Do not leave the assignment of an instructor in the mathematics appreciation course until the last minute and do not assign it on the sole basis of availability. The course requires more planning and preparation than almost any other mathematics course if it is to be successful.

2. Do not simply allow the students to sit back and listen. It is important that they be involved actively. But this need not take the form of daily homework. In fact, drill type assignments should be avoided. The involvement could take the form of projects, papers, book reports, "discovering" mathematics in class, participating in class discussions.

3. Do not overemphasize the history of mathematics. While the history of mathematics could and should be used to enliven the topics covered, a student who knows (and cares) nothing about a mathematical topic is not likely to be interested in its history.

4. Do not stress remedial topics. While many of the students in a mathematics appreciation course may need remedial work, any such material that is covered must be presented as part of a topic that fits into the scope of the course as a whole.

5. Do not make a fetish of rigor; in particular do not prove things that are self-evident to the students. For example, a rigorous presentation of the real numbers in which one proves the uniqueness of zero is entirely inappropriate in courses of this type.

6. Do not cover topics you do not yourself find interesting and important. It is hard to fool these students, and if the teacher does not care, they will not see why they should.

7. Do not be condescending. While the students in such courses may not be mathematically inclined, this does not mean that they are unintelligent. Many who take mathematics appreciation courses are outstanding, creative students, who have simply concentrated in the nonquantitative areas of the curriculum. The attitude of the teacher can help either to open or to close their minds to the material.

8. Do not cover topics which you cannot relate in some way to ideas familiar to the students.

Clock arithmetic and symbolic logic, for example, are of little value to mathematics appreciation courses unless you can find applications the students can appreciate and understand.

9. Do not make the course too easy. The material should not be way over the heads of the students, but it should not be trivial either.

10. Do not accept anyone else's blueprint for a mathematics appreciation course. If you can communicate, in your own way, why you believe that mathematics is beautiful and important, the course will fulfill its purpose.

IV. Approaches to Course Organization. There are nearly as many ways to teach a course in mathematics appreciation as there are teachers of these courses. While some strategies will work superbly in some contexts, none can be recommended for all; the teacher's enthusiasm for what is being done as well as the appropriateness of the strategy for the students in the course are generally more important than the actual strategy adopted. Nevertheless, to encourage flexibility, we list below some approaches to teaching mathematics appreciation that have been effective in certain contexts.

1. A sampler approach, featuring a variety of more or less independent topics. The advantage of this method is that it covers many areas without requiring a sustained continuity of interest; students who fall behind or simply fail to comprehend one topic always know that they have a chance for a fresh start in a few days. The disadvantage is that of all survey courses: not enough time spent on any one thing to ensure long-term learning.

2. A single-thread approach, built around a common theme, for example, 2×2 matrices, or algorithms, or patterns of symmetry. Doing this takes careful planning, and runs the risk of alienating some of the class who find the thread incomprehensible. But it guarantees a solid example of the intellectual coherence that is so much a part of contemporary mathematics, that ideas arising in one context find applications in others, and that a common abstract structure underlies them all.

3. A socratic approach, in which the instructor works carefully to let the students develop their own reasoning. This works well in small classes with a highly motivated instructor. While the content of such courses is hard to guarantee in advance, the achievement for students who are able to think for themselves, perceiving patterns where others simply see chaos, is a worthy objective for a course in mathematics appreciation.

V. Examples of Topics. The topics available for courses in mathematics appreciation are as diverse as mathematical science itself. Standard textbooks offer a rather traditional assortment of topics: probability, graph theory, finite difference equations, computers, matrices, statistics, exponential growth, set theory, and logic seem to dominate. But there are numerous other themes that can be used for large or small components of courses. Here are a few of the many possible examples.

1. Understanding how to use the buttons on a pocket calculator. It used to be that the number e was a complete mystery to those who had not studied calculus, and that "sin" had for humanities students more the connotation of theology than of mathematics. But no more. Virtually everyone has, or has seen, inexpensive hand calculators with buttons that perform operations involving exponential, trigonometric, and basic statistical functions. Teaching a class what these buttons do is an exciting new way to explore some traditional parts of classical mathematics.

2. Tracing the modern descendants of classical mathematical ideas can illustrate the power of mathematics to influence the real world, as well as its remoteness from it. For example, classical Greek geometry involving conic sections led to models for planetary motion, and ultimately to the

possibility of space flight. And probability, which had its origins in seventeenth-century discussions about gambling, now dominates actuarial and fiscal policy, influencing government and corporate budgets, thus affecting the level of interest, of unemployment, and the health of the entire economy.

3. Connecting mathematics with Nobel Prizes. Nobel prizes are not given in mathematics (and the apocryphal reasons for this are quite amusing). But the work that led to Nobel prizes (e.g., of Libby, of Arrow, of Lederberg, and others) often has an intrinsically mathematical basis. The study of this scientific work provides an opportunity to show how mathematics is important in the most profound discoveries of modern science.

4. Applying exponential growth models. The applications of traditional topics from elementary mathematics can often be explored more fully than has usually been the case. Exponential growth and decay models provide a striking example. Simple noncalculus approaches to models of growth provide a basis for discussion not only of interest and inflation, but also of such things as radiocarbon dating, cooling and heating of houses, population dynamics, strategies for controlling epidemics, and even detection of art forgeries.

5. Relating traditional mathematics to new applications. A discussion of beginning probability theory can quickly lead to a treatment of the Hardy-Weinberg law of genetics and a calculation of the probability of winning state lotteries. An introductory treatment of statistics can quickly lead to a discussion of political polls, the design and interpretation of surveys and of related decision-making problems. Modern applications of elementary network theory include recent work in computational complexity and almost unbreakable codes.

6. Introducing problems involving decision-making. There are many situations described by elementary mathematics in which one must choose "rationally" among possible options. One can discuss quantifying risk and uncertainty, fair division schemes, applications of network flows, pursuit and navigation problems, game theory and numerous other topics. Political science is full of unexpected but usually interesting topics, including Arrow's theorem and its offshoot theories of voting, the recently discovered problems associated with apportionment of legislatures, and strategies of fair voting in multiple candidate elections.

7. Exploring the powers and limitations of mathematical models. Each of the modern social sciences abounds with applications of elementary mathematics. All of the examples mentioned above, and many more, involve the use of mathematical models. Sometimes these models are quite accurate and sometimes they are not. But even in the latter case the model can help clarify one's thinking about the underlying problem. An example of this use of mathematical modelling is the prisoner's dilemma argument of game theory and its possible connection with U.S.-U.S.S.R. relations.

References

Numerous references for mathematics appreciation courses are given in Section IX of this report. We list here only a few specific suggestions for the numbered topics mentioned above.

1. See the handbooks for various calculators.

2. Much of this is in standard textbooks. Morris Kline's *Mathematics in Western Culture* and George Pólya's *Mathematical Models in Science* are helpful sources.

3. Libby's work is briefly discussed in several elementary texts on ordinary differential equations. e.g., in *Differential Equations with Applications and Historical Notes* by George F. Simmons. For some work of Arrow see Edward Bender, *An Introduction to Mathematical Modeling*, Wiley, New York, 1978. Lederberg published an article in this MONTHLY entitled "Hamilton circuits of convex trivalent polyhedra (up to 18 vertices)" in vol. 74, pp. 522–527.

4. See any modern text on ordinary differential equations. A particularly good one is Martin Braun, *Differential Equations and Their Applications*, Springer-Verlag, New York, 1975.

5. The Hardy-Weinberg law appears in several texts on finite mathematics, e.g., *Applied Finite Mathematics* by Anton and Kolman. *How to Lie with Statistics* by Darrel Huff and Irving Geis, Norton, 1954, and other texts contain situational mathematics which can be discussed according to the interests of the audience. For the two topics mentioned last, see *Scientific American*, Jan. 1978, p. 96, and Aug. 1977.

6. See Bender (as in 3); the articles by William Lucas in vol. 2 of the forthcoming *Modules in Applied Mathematics* (Springer-Verlag, New York); M. Balinski and H. P. Young, *Proc. Nat. Acad. Sci. U.S.A.*, 77 (January 1980) 1–4; H. Hamburger, *J. Math. Sociology*, 3 (1973) 27–48; David Gale, UMAP Module 317, 1978; W. Stromquist, this MONTHLY 87 (1980) 640–644; and George Minty's article in M. D. Thompson, ed., *Discrete Mathematics and its Applications* (Indiana University, Bloomington, 1977).

VI. Two-Year Colleges. Many courses that ought to follow the "mathematics appreciation" philosophy are taught in two-year colleges. Innovative approaches and curriculum development by some two-year faculty are reflected by their texts and articles in this area. Although the preceding sections of this report are applicable to mathematics appreciation courses in all colleges, this separate section appears because of the special problems created in two-year colleges by generally heavy teaching loads, by staffing in some cases by faculty whose mathematical experiences are not sufficient to make them comfortable with the broad range of topics demanded by these courses, and by the regrettable frequency of administrative procedures which allow students needing remediation to enroll in these courses. The following suggestions may help to overcome these impediments to two-year college implementation of the goals of mathematics appreciation courses.

1. When there is a choice among faculty members for assignment to the mathematics appreciation course, only those having a broad range of mathematical experiences and expressing interest in the course should teach it. Mathematics program administrators should provide extra guidance to faculty teaching this course for the first time. In the two-year college there will usually be one text used by all teachers, often supplemented by a reading list and/or other texts; a description of special uses of these materials, as well as sample course outlines, supplementary and classwork materials, and tests, will be helpful. Entrance and/or exit requirements may be matters of policy and should be explained. Lists of applicable resource material owned by the school should be provided to new teachers, along with knowledge of the school's film rental policies.

2. Special attention should be paid to the needs of this course by the library, audio-visual, and computer facilities. The mathematics program administrator should be sure these courses are adequately supported.

3. Since mathematics appreciation courses, properly taught, take an enormous amount of preparation time, any load relief possible would be appropriate. In a suitable lecture room, the course can be effectively taught to "double" sections of 60–90 students if doing so would leave the teacher several hours more preparation time each week. (Such a load might be counted as two or three sections, corresponding to the grading load.)

4. Sharing materials and ideas and perhaps team-teaching would be reasonable for mathematics appreciation courses. One teacher at a school might be most qualified for teaching, say, a computer unit, and might "rotate" across several sections. Many more "hand-out" materials seem to be necessary for mathematics appreciation courses than for traditional courses; these might be used by several teachers in a given term, or re-used in succeeding terms. Faculty teaching mathematics appreciation courses seem to enjoy sharing materials and methods.

5. In-course remediation should be avoided. If students are enrolled who cannot handle elementary operations at the level needed for the work of the course, a "math lab" facility might

be used to design and administer individual remediation programs. It cannot be over-emphasized that a mathematics appreciation course cannot fulfill its goal if it degenerates into the teaching of arithmetic computations or pre-algebra skills, or if it is limited to a topic such as "consumer mathematics."

6. A large proportion of students enter two-year colleges with little realistic expectation concerning majors. Many of these students have had poor experiences with mathematics and, if there is a general education mathematics requirement which may be satisfied by either a mathematics appreciation course or a pre-calculus/calculus course, they will often elect the mathematics appreciation course. Well into a successful term, the student may begin to think realistically about mathematics requirements of various university majors. Since most majors outside the humanities will necessitate at least some mathematics at a technical level rarely achieved in the typical two-year college mathematics appreciation course, an important service of this course can be to channel these students back into regular sequence mathematics courses. Without violating the spirit of a mathematics appreciation course, it is possible to include a topically organized unit requiring the review and use of elementary algebra and graphing techniques; this may give the student a successful experience in doing mathematics that serve as encouragement to return to regular sequence mathematics courses. (A linear programming unit, for example, requires the students to review or acquire facility with graphing and algebra techniques. Many of the topics suitable for a mathematics appreciation course can be handled in this way.) Students with the experience will frequently place higher in the sequence courses than they would have upon original enrollment, and will go on as solid, though late-blooming, students.

VII. Films. Since students in mathematics appreciation courses frequently have little experience in sustaining interest in regular mathematics lectures, it is usually appropriate in these courses to provide a variety of class activities. Films are a useful but under-utilized medium for mathematics instruction generally. They are especially useful for the mathematics appreciation course.

We list in the Center Section a selection of films about mathematical subjects that are suitable for lay audiences. (Distributor addresses are listed at the end.) Further information on these and other films is available in the booklet *Annotated Bibliography of Films and Videotapes for College Mathematics* by David Schneider (M.A.A., 1980).

VIII. Classroom Aids. Certain topics treated in mathematics appreciation courses are particularly amenable to demonstration with physical or geometric devices. Useful exhibits can often be seen at NCTM meetings. A list of major suppliers of mathematics classroom devices is given in the Center Section.

IX. References. Since many of the topics that arise in mathematics appreciation courses occur nowhere else in the mathematics curriculum, it is quite important that instructors be aware of the expository literature of mathematics that treats its relations to science and society. Student term papers in courses on mathematics appreciation typically tax the instructor's knowledge of the literature more than any other course in the mathematics curriculum.

To aid instructors of mathematics appreciation courses, we list in the Center Section major references that would be suitable for background reading, and as sources for special projects. This list does not include textbooks, partly because we do not wish to endorse some books over others, and partly because texts go in and out of print much more rapidly than the reference classics.

ANSWER TO "PHOTO" ON PAGE 11

Julia Robinson, current president of the American Mathematical Society. Photo by George Bergman.

A MOCK SYMPOSIUM FOR YOUR CALCULUS CLASS

DENNIS WILDFOGEL
Mathematics Program, Stockton State College, Pomona, NJ 08240

Dr. Gregory Campe, Professor of Organic Chemistry at Frostbite Falls (Minn.) State College. Dr. Linda Gillespe, Professor of Environmental Science at the Susan B. Anthony University for Women. Dr. Daniel Hain, Professor of Surfing at Could's Hole Oceanographic Institute. An unknowing colleague of mine at Stockton State College looked at the impressive roster of thirty-five participants in the First Stockton Symposium on Mathematical Modeling in the Life Sciences and exclaimed, "You got all those people to come here?" What he didn't know was that the people on the roster were the students in my Calculus For Life Scientists II class, given phony titles and affiliations at fictitious institutes.

Calculus For Life Scientists is a two-semester, introductory level calculus course aimed primarily at students majoring in environmental science, marine science, and biology. The course emphasizes mathematical modeling. At the "Symposium," which acts as a capstone to the students' year-long encounter with calculus, teams of students make presentations about a modeling problem on which they have been working for about two weeks. The benefits of doing these projects are numerous: (1) the projects give the students a holistic view of the use of higher mathematics in their own disciplines; (2) most projects require them to learn some techniques they would not ordinarily encounter in a one-year calculus course; (3) they learn about the benefits and difficulties of working on a team under a pressing deadline; (4) they experience the gamut of emotions associated with the problem-solving process, from the pleasure of initial idea generation, through the frustration of the intermediate stages, to the triumph of the completed project; (5) they gain experience in making oral presentations and written reports; (6) they see how the need for mathematical techniques grows out of realistic problems.

I have conducted such a symposium four times now. Students consistently rate it as one of their most rewarding and useful academic experiences. Several other faculty members have successfully adapted the symposium idea for use in their own classes.

Here's how it works. About two and a half weeks before the end of the term, I divide the class of thirty to forty students into teams of about six students each. Each team is given a fairly difficult modeling problem and is responsible for making both an oral presentation and a written report about the results of their investigation of that problem. Each team meets during class time for the remainder of the term while I give hints and feedback on their work. The teams invariably find it necessary to schedule meetings outside of class time, too. I always offer enough suggestions so that each team will develop a good model by the time of the symposium.

The key aspect of running a symposium like this is the selection of appropriate problems for the teams. Each problem must be sufficiently challenging to occupy a team of six students for two weeks and yet still be within reach of their capabilities. There are a few texts ([**1**]-[**3**]) which I have used repeatedly as sources for problems. In a few instances I have been able to adapt material from books or journals in other fields, and I have made up several problems on my own. Colleagues in the life sciences have provided a great deal of assistance. Below is a partial list of titles of problems I have used.

Competition of two species for limited resources
Biogeography: a species equilibrium model
The maximum brightness of Venus
Operating strategies for publicly owned commuter bus systems
A box model for airshed pollutant capacity estimation
The effects of natural selection on gene frequency
Passive transport of chemical substances through a thick section of tissue
Determination of the shape of subterranean deposits by use of gravitational anomalies

Parasitic relationships which are not harmful to the host
Ventilation systems and the accumulation of toxic pollutants
A model for the clinical detection of diabetes
A rare example of a closed ecological system
The chemical kinetics of bimolecular reactions
Excretion of a drug
An optimal inventory policy model for an import wholesaler

The composition of the student teams is important. I make sure that each student works on a project in an area of his or her own interest. I have tried dividing the class into groups homogeneous or heterogeneous according to ability. The homogeneous groups make it easier to tailor the difficulty of the problem to the appropriate level; however, it is difficult to keep the least capable groups from becoming discouraged. In heterogeneous groupings, the less successful students can learn from the better ones; however, too much of the burden then falls on the better students in each group. The best compromise I have found so far is to have two "all-star" teams of the best students, and to have heterogeneous teams composed of the remaining students.

The symposium itself occupies the last three days of the term. Each team makes a half-hour presentation. The entire event, always attended by several other faculty members, is done up in tongue-in-cheek style. I circulate in advance to all mathematics and science faculty members a Roster of Participants and a Schedule of the Symposium, making it look as much like a real scholarly meeting as possible. (I always fool at least one new faculty member!)

At the beginning of each session, one of my mathematics or science colleagues makes a humorous presentation, e.g., a "double talk" address that sounded like a commentary on the specific models to be presented, or a short discourse on the three-body problem while juggling three balls. Once Miss America visited, and another time I sang a song about calculus which I composed. The student teams get into the act, dressing up in suits and ties or lab coats, calling each other "Doctor" or "Professor," and occasionally putting on brief skits. The merriment makes it enjoyable without detracting from the serious work to be done and serves to alleviate some of the anxiety the students have about making oral presentations. Afterwards, each student receives a copy of the "Proceedings of the Symposium" containing the written reports of the several teams and a few memorable photographs of the symposium.

The symposium creates an opportunity for students to understand the way mathematics is actually used in their own fields and to understand both their own potential as users of mathematics and the difficulties inherent in the modeling process. It is thus a valuable experience in their mathematical education.

References

1. M. Braun, Differential Equations and Their Applications: An Introduction to Applied Mathematics, 2nd ed., Springer-Verlag, New York, 1978.

2. D. P. Maki and M. Thompson, Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences, Prentice-Hall, Englewood Cliffs, N. J., 1973.

3. E. O. Wilson and W. H. Bossert, A Primer of Population Biology, Sinauer Associates, Stamford, Conn., 1971.

MISCELLANEA

90.

"Throughout these works, grasping the thought in *P* in the sense of *having* the thought that *P* appears to be confused with grasping the thought that *P* in the sense of having the thought that a certain *other* thought *is* the thought that *P*."

— Gilbert Harman, in a review in the *Times Literary Supplement*, 16 April 1982.

Parasitic relationships which are not harmful to the host
Ventilation systems and the accumulation of toxic pollutants
A model for the clinical detection of diabetes
A rare example of a closed ecological system
The chemical kinetics of bimolecular reactions
Excretion of a drug
An optimal inventory policy model for an import wholesaler

The composition of the student teams is important. I make sure that each student works on a project in an area of his or her own interest. I have tried dividing the class into groups homogeneous or heterogeneous according to ability. The homogeneous groups make it easier to tailor the difficulty of the problem to the appropriate level; however, it is difficult to keep the least capable groups from becoming discouraged. In heterogeneous groupings, the less successful students can learn from the better ones; however, too much of the burden then falls on the better students in each group. The best compromise I have found so far is to have two "all-star" teams of the best students, and to have heterogeneous teams composed of the remaining students.

The symposium itself occupies the last three days of the term. Each team makes a half-hour presentation. The entire event, always attended by several other faculty members, is done up in tongue-in-cheek style. I circulate in advance to all mathematics and science faculty members a Roster of Participants and a Schedule of the Symposium, making it look as much like a real scholarly meeting as possible. (I always fool at least one new faculty member!)

At the beginning of each session, one of my mathematics or science colleagues makes a humorous presentation, e.g., a "double talk" address that sounded like a commentary on the specific models to be presented, or a short discourse on the three-body problem while juggling three balls. Once Miss America visited, and another time I sang a song about calculus which I composed. The student teams get into the act, dressing up in suits and ties or lab coats, calling each other "Doctor" or "Professor," and occasionally putting on brief skits. The merriment makes it enjoyable without detracting from the serious work to be done and serves to alleviate some of the anxiety the students have about making oral presentations. Afterwards, each student receives a copy of the "Proceedings of the Symposium" containing the written reports of the several teams and a few memorable photographs of the symposium.

The symposium creates an opportunity for students to understand the way mathematics is actually used in their own fields and to understand both their own potential as users of mathematics and the difficulties inherent in the modeling process. It is thus a valuable experience in their mathematical education.

References

1. M. Braun, Differential Equations and Their Applications: An Introduction to Applied Mathematics, 2nd ed., Springer-Verlag, New York, 1978.

2. D. P. Maki and M. Thompson, Mathematical Models and Applications: With Emphasis on the Social, Life, and Management Sciences, Prentice-Hall, Englewood Cliffs, N. J., 1973.

3. E. O. Wilson and W. H. Bossert, A Primer of Population Biology, Sinauer Associates, Stamford, Conn., 1971.

MISCELLANEA

90.

"Throughout these works, grasping the thought in *P* in the sense of *having* the thought that *P* appears to be confused with grasping the thought that *P* in the sense of having the thought that a certain *other* thought *is* the thought that *P*."

—Gilbert Harman, in a review in the *Times Literary Supplement*, 16 April 1982.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by May 31, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2980. *Proposed by Jordi Dou, Barcelona, Spain.*

Given the points A_1, A_2, A_3, M and the line s, construct P, Q such that $\overline{PQ}$ is equal and parallel to $\overline{A_1M}$ and $\overline{P_1Q_1} = \overline{P_2Q_2} = \overline{P_3Q_3}$, where P_i, Q_i are the intersections of PA_i, QA_i with s.

Describe the locus of the point M for which the problem has a solution when A_1, A_2, A_3 and s are known (fixed).

E 2981. *Proposed by Murray S. Klamkin, University of Alberta, Canada.*

If the three medians of spherical triangle are equal, must the triangle be equilateral? Note that the sides of a (proper) spherical triangle are *minor* arcs of great circles and thus its perimeter is $< 2\pi$.

E 2982. *Proposed by R. L. Graham, Bell Telephone Laboratories, and D. E. Knuth, Stanford University.*

Let $\|y\|$ denote the distance from (real) y to the nearest integer. Evaluate the double infinite sum

$$\cdots + 4\|x/4\|^2 + 2\|x/2\|^2 + \|x\|^2 + \|2x\|^2/2 + \|4x\|^2/4 + \cdots.$$

E 2983. *Proposed by E. Ehrhart, University of Strasbourg, France.*

Let ABC be an equilateral triangle of perimeter $3a$. Calculate the area of the convex region consisting of all points P such that $PA + PB + PC \leqslant 2a$.

E 2984. *Proposed by Roger Maddux, Iowa State University.*

Characterize the logically valid sentences of the form

$$Q_1x_1 \dots Q_nx_nRx_1 \dots x_n \to Q_1'x_{i_1} \dots Q_n'x_{i_n}Rx_1 \dots x_n,$$

where Q_i is $\forall$ or $\exists$, $i_1 \dots i_n$ is a permutation of $1 \dots n$, and R is a relation.

E 2985. *Proposed by P. J. Forrester, University of Melbourne and M. L. Glasser, Clarkson College.*

Show that for $0 < x < 1$

$$\prod_{\substack{k=1\\ \text{odd}}}^{\infty} \left[(1+x^k)/(1-x^k)\right]^{\phi(k)/k} = \exp\left[2/(x^{-1}-x)\right]$$

where $\phi(k)$ is Euler's totient.

SOLUTIONS OF ELEMENTARY PROBLEMS

Equivalence of Two Triangles

E 2727* [1978, 594; 1979, 790]. *Proposed by David P. Robbins, Hamilton College.*

Two triangles $A_1A_2A_3$ and $B_1B_2B_3$ in $\mathbf{R}^3$ are equivalent if there exist three different parallel lines p_1, p_2, p_3 and rigid motions σ, τ such that $\sigma(A_i)$ and $\tau(B_i)$ lie on p_i $(i = 1,2,3)$.

Find necessary and sufficient conditions for equivalence of two triangles.

Solution by M. J. Pelling, Balliol College, Oxford. Let the triangles $A_1A_2A_3$, $B_1B_2B_3$ have respective sides a, b, c and d, e, f. We shall show that $A_1A_2A_3$, $B_1B_2B_3$ are equivalent unless either

$$\sqrt{a^2-d^2}, \sqrt{b^2-e^2}, \sqrt{c^2-f^2} \quad \text{or} \quad \sqrt{d^2-a^2}, \sqrt{e^2-b^2}, \sqrt{f^2-c^2}$$

form the sides of a proper triangle. From consideration of the normal section with sides α, β, γ of the triangular prism bounded by parallel lines p_1, p_2, p_3 and from the published solution [1979, 790] we see the triangles are equivalent if and only if $\alpha, \beta, \gamma > 0$ can be found forming the sides of a proper triangle, with $\alpha^2 \le \min(a^2, d^2)$ etc. and

$$\pm\sqrt{a^2-\alpha^2} \pm \sqrt{b^2-\beta^2} \pm \sqrt{c^2-\gamma^2} = 0$$

$$\pm\sqrt{d^2-\alpha^2} \pm \sqrt{e^2-\beta^2} \pm \sqrt{f^2-\gamma^2} = 0. \tag{1}$$

Now the surface $\pm\sqrt{x} \pm \sqrt{y} \pm \sqrt{z} = 0$ is the right circular cone

$$\Gamma\colon x^2+y^2+z^2-2yz-2zx-2xy=0$$

with axis in direction $(1,1,1)$, vertex O and semiangle $\tan^{-1}\frac{1}{2}\sqrt{2}$. One half-cone Γ_+ lies within the octant $x, y, z \ge 0$ and the other half-cone Γ_- lies in the octant $x, y, z \le 0$. If $\operatorname{Int}\Gamma_+$ denotes the open region bounded by Γ_+, then our key observation is that $(x, y, z) \in \operatorname{Int}\Gamma_+$ if and only if $\sqrt{x}, \sqrt{y}, \sqrt{z}$ are the sides of a proper triangle. Denote by $\Gamma(p, q, r)$ the cone Γ translated to have vertex at (p, q, r) with $\Gamma_\pm(p, q, r)$ analogously defined.

The conditions on α, β, γ are satisfied if and only if

$$\Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2) \cap \operatorname{Int}\Gamma_+ \neq \varnothing \tag{2}$$

and we shall show that if $\Gamma_-(a^2, b^2, c^2)$ and $\Gamma_-(d^2, e^2, f^2)$ intersect at all, then already they have nonempty intersection within $\operatorname{Int}\Gamma_+$.

Let π_0 be the plane through O normal to the axis of Γ and let the axes of $\Gamma_-(a^2, b^2, c^2)$, $\Gamma_-(d^2, e^2, f^2)$ meet π_0 in O_1, O_2 respectively. These two half-cones cut π_0 in respective circles C_1 centre O_1 radius $R_1 > 0$ and C_2 centre O_2 radius $R_2 > 0$, and

$$\Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2) \neq \varnothing$$

if and only if $|R_2 - R_1| \le O_1O_2$. The parallel plane π_t at distance $t\sqrt{2}$ from π_0 (on the Γ_+ side) cuts $\Gamma_+, \Gamma_-(a^2, b^2, c^2)$, $\Gamma_-(d^2, e^2, f^2)$ in 3 circles which project orthogonally onto π_0 into 3 circles $C(t), C_1(t), C_2(t)$ with respective centres O, O_1, O_2 and radii $t, R_1 - t, R_2 - t$. The part of

E 2985. *Proposed by P. J. Forrester, University of Melbourne and M. L. Glasser, Clarkson College.*

Show that for $0 < x < 1$

$$\prod_{\substack{k=1 \\ \text{odd}}}^{\infty} \left[(1 + x^k)/(1 - x^k)\right]^{\phi(k)/k} = \exp\left[2/(x^{-1} - x)\right]$$

where $\phi(k)$ is Euler's totient.

SOLUTIONS OF ELEMENTARY PROBLEMS

Equivalence of Two Triangles

E 2727* [1978, 594; 1979, 790]. *Proposed by David P. Robbins, Hamilton College.*

Two triangles $A_1A_2A_3$ and $B_1B_2B_3$ in $\mathbf{R}^3$ are equivalent if there exist three different parallel lines p_1, p_2, p_3 and rigid motions σ, τ such that $\sigma(A_i)$ and $\tau(B_i)$ lie on p_i ($i = 1, 2, 3$).

Find necessary and sufficient conditions for equivalence of two triangles.

Solution by M. J. Pelling, Balliol College, Oxford. Let the triangles $A_1A_2A_3$, $B_1B_2B_3$ have respective sides a, b, c and d, e, f. We shall show that $A_1A_2A_3$, $B_1B_2B_3$ are equivalent unless either

$$\sqrt{a^2 - d^2}\,, \sqrt{b^2 - e^2}\,, \sqrt{c^2 - f^2} \quad \text{or} \quad \sqrt{d^2 - a^2}\,, \sqrt{e^2 - b^2}\,, \sqrt{f^2 - c^2}$$

form the sides of a proper triangle. From consideration of the normal section with sides α, β, γ of the triangular prism bounded by parallel lines p_1, p_2, p_3 and from the published solution [1979, 790] we see the triangles are equivalent if and only if $\alpha, \beta, \gamma > 0$ can be found forming the sides of a proper triangle, with $\alpha^2 \le \min(a^2, d^2)$ etc. and

$$\pm\sqrt{a^2 - \alpha^2} \pm \sqrt{b^2 - \beta^2} \pm \sqrt{c^2 - \gamma^2} = 0$$

$$(1) \qquad \pm\sqrt{d^2 - \alpha^2} \pm \sqrt{e^2 - \beta^2} \pm \sqrt{f^2 - \gamma^2} = 0.$$

Now the surface $\pm\sqrt{x} \pm \sqrt{y} \pm \sqrt{z} = 0$ is the right circular cone

$$\Gamma\colon x^2 + y^2 + z^2 - 2yz - 2zx - 2xy = 0$$

with axis in direction (1, 1, 1), vertex O and semiangle $\tan^{-1}\frac{1}{2}\sqrt{2}$. One half-cone Γ_+ lies within the octant $x, y, z \ge 0$ and the other half-cone Γ_- lies in the octant $x, y, z \le 0$. If $\operatorname{Int}\Gamma_+$ denotes the open region bounded by Γ_+, then our key observation is that $(x, y, z) \in \operatorname{Int}\Gamma_+$ if and only if $\sqrt{x}, \sqrt{y}, \sqrt{z}$ are the sides of a proper triangle. Denote by $\Gamma(p, q, r)$ the cone Γ translated to have vertex at (p, q, r) with $\Gamma_\pm(p, q, r)$ analogously defined.

The conditions on α, β, γ are satisfied if and only if

$$(2) \qquad \Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2) \cap \operatorname{Int}\Gamma_+ \neq \varnothing$$

and we shall show that if $\Gamma_-(a^2, b^2, c^2)$ and $\Gamma_-(d^2, e^2, f^2)$ intersect at all, then already they have nonempty intersection within $\operatorname{Int}\Gamma_+$.

Let π_0 be the plane through O normal to the axis of Γ and let the axes of $\Gamma_-(a^2, b^2, c^2)$, $\Gamma_-(d^2, e^2, f^2)$ meet π_0 in O_1, O_2 respectively. These two half-cones cut π_0 in respective circles C_1 centre O_1 radius $R_1 > 0$ and C_2 centre O_2 radius $R_2 > 0$, and

$$\Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2) \neq \varnothing$$

if and only if $|R_2 - R_1| \le O_1O_2$. The parallel plane π_t at distance $t\sqrt{2}$ from π_0 (on the Γ_+ side) cuts $\Gamma_+, \Gamma_-(a^2, b^2, c^2)$, $\Gamma_-(d^2, e^2, f^2)$ in 3 circles which project orthogonally onto π_0 into 3 circles $C(t)$, $C_1(t)$, $C_2(t)$ with respective centres O, O_1, O_2 and radii t, $R_1 - t$, $R_2 - t$. The part of

the curve $\Gamma_-(a^2,b^2,c^2)\cap\Gamma_-(d^2,e^2,f^2)$ on the Γ_+ side of π_0 projects into the locus of $C_1(t)\cap C_2(t)$ as t varies between 0 and $\min(R_1,R_2)$, and this locus is the portion H of the hyperbola in π_0 defined by $O_2P-O_1P=R_2-R_1$ (assuming $R_2\geq R_1$) for $O_1P\leq R_1$, $O_2P\leq R_2$. We have to show there is some value of t for which one of the points P_t of $C_1(t)\cap C_2(t)$ exists and is inside the circle $C(t)$, i.e., $OP_t<t$.

Now $O\in \operatorname{Int} C_1\cap \operatorname{Int} C_2$ since $(a^2,b^2,c^2),(d^2,e^2,f^2)\in \operatorname{Int}\Gamma_+$ (which implies $C_1(0)\cap C_2(0)$ contains 2 points and H is not empty) and as t increases from 0 there is a first value $t^*>0$ for which $C(t^*)$ just touches $C_1(t^*)$ or $C_2(t^*)$, say $C_1(t^*)$ at Q. Let H^* denote the portion of H defined by $C_1(t)\cap C_2(t)$ for $t\geq t^*$: then H^* must meet O_1Q internally (draw a diagram, noting that $C(t^*)$ lies inside both $C_1(t^*)$ and $C_2(t^*)$) in a point P_t with $t>t^*$ and

$$OP_t=|OQ-(t-t^*)|=|2t^*-t|<t.$$

It follows that (2) is equivalent simply to

$$\Gamma_-(a^2,b^2,c^2)\cap\Gamma_-(d^2,e^2,f^2)\neq\varnothing \tag{3}$$

and this holds if and only if $(a^2,b^2,c^2)\notin \operatorname{Int}\Gamma_+(d^2,e^2,f^2)$ and $(d^2,e^2,f^2)\notin \operatorname{Int}\Gamma_+(a^2,b^2,c^2)$, which is equivalent to

$$(a^2-d^2,b^2-e^2,c^2-f^2),(d^2-a^2,e^2-b^2,f^2-c^2)\notin \operatorname{Int}\Gamma_+.$$

So $A_1A_2A_3$, $B_1B_2B_3$ are equivalent unless either

$$\sqrt{a^2-d^2},\sqrt{b^2-e^2},\sqrt{c^2-f^2}\quad\text{or}\quad\sqrt{d^2-a^2},\sqrt{e^2-b^2},\sqrt{f^2-c^2}$$

form the sides of a proper triangle.

A Third Degree Congruence with Nine Solutions

E 2763 [1979, 223]. *Proposed by Lorraine L. Foster, California State University, Northridge.*

Let $f(n)=n^3+396n^2-111n+38$. Prove that the congruence $f(n)\equiv 0 \pmod{3^a}$ has precisely 9 solutions $(\operatorname{mod} 3^a)$ for all integers $a\geqslant 5$.

Solution by L. E. Mattics, University of South Alabama.

THEOREM. *If $s\geq 3$, then there is a d such that $f(d)\equiv 0\ (3^{s+2})$ and if $f(b)\equiv 0\ (3^{s+2})$ then $b\equiv d$ $(\operatorname{mod} 3^s)$.*

Indeed, $f(n)\equiv 0 \pmod{243}$ if and only if $n\equiv 19 \pmod{27}$ and we proceed by induction assuming the theorem to be true for some $s\geq 3$. Then

$$f(d+d3^s)\equiv f(d)+k3^sf'(d) \pmod{3^{s+3}}$$

and since $d\equiv 1 \pmod 9$, we can show that $3^2|f'(d)$, but $3^3\nmid f'(d)$. Hence the congruence

$$f(d)+k3^sf'(d)\equiv 0 \pmod{3^{s+1}}$$

can be solved uniquely for k modulo 3. If $f(c)\equiv 0 \pmod{3^{s+2}}$, then by induction hypothesis $c=a+d\cdot 3^s$ so $d\equiv k \pmod 3$, and $c\equiv a+k3^s \pmod{3^{s+1}}$. The induction is complete and the theorem is proved.

Now if $s\geq 5$ and $f(b)\equiv 0 \pmod{3^s}$, then

$$f(b+m3^{s-2}+r3^{s-1})\equiv f(b)+m3^{s-2}f'(b)+r3^{s-1}f'(b+m\cdot 3^{s-2}),$$

and this quantity is $\equiv 0 \pmod{3^s}$ because $3^2|f'(b)$ and $3|f'(b+m3^{s-2})$. Therefore, since b is unique modulo 3^{s-2}, there are precisely nine solutions modulo 3^s to the congruence $f(n)\equiv 0\ (3^s)$.

Also solved by H. L. Abbott (Canada), J. Ampe (student), L. Kuipers (Switzerland), Jinku Lee (student), N. Miku (Netherlands), A. M. Rockett, R. S. Stacy, P. J. Zwier, and the proposer.

the curve $\Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2)$ on the Γ_+ side of π_0 projects into the locus of $C_1(t) \cap C_2(t)$ as t varies between 0 and $\min(R_1, R_2)$, and this locus is the portion H of the hyperbola in π_0 defined by $O_2P - O_1P = R_2 - R_1$ (assuming $R_2 \geq R_1$) for $O_1P \leq R_1, O_2P \leq R_2$. We have to show there is some value of t for which one of the points P_t of $C_1(t) \cap C_2(t)$ exists and is inside the circle $C(t)$, i.e., $OP_t < t$.

Now $O \in \operatorname{Int} C_1 \cap \operatorname{Int} C_2$ since $(a^2, b^2, c^2), (d^2, e^2, f^2) \in \operatorname{Int} \Gamma_+$ (which implies $C_1(0) \cap C_2(0)$ contains 2 points and H is not empty) and as t increases from 0 there is a first value $t^* > 0$ for which $C(t^*)$ just touches $C_1(t^*)$ or $C_2(t^*)$, say $C_1(t^*)$ at Q. Let H^* denote the portion of H defined by $C_1(t) \cap C_2(t)$ for $t \geq t^*$: then H^* must meet O_1Q internally (draw a diagram, noting that $C(t^*)$ lies inside both $C_1(t^*)$ and $C_2(t^*)$) in a point P_t with $t > t^*$ and

$$OP_t = |OQ - (t - t^*)| = |2t^* - t| < t.$$

It follows that (2) is equivalent simply to

$$\Gamma_-(a^2, b^2, c^2) \cap \Gamma_-(d^2, e^2, f^2) \neq \varnothing \tag{3}$$

and this holds if and only if $(a^2, b^2, c^2) \notin \operatorname{Int} \Gamma_+(d^2, e^2, f^2)$ and $(d^2, e^2, f^2) \notin \operatorname{Int} \Gamma_+(a^2, b^2, c^2)$, which is equivalent to

$$(a^2 - d^2, b^2 - e^2, c^2 - f^2), (d^2 - a^2, e^2 - b^2, f^2 - c^2) \notin \operatorname{Int} \Gamma_+.$$

So $A_1A_2A_3$, $B_1B_2B_3$ are equivalent unless either

$$\sqrt{a^2 - d^2}, \sqrt{b^2 - e^2}, \sqrt{c^2 - f^2} \quad \text{or} \quad \sqrt{d^2 - a^2}, \sqrt{e^2 - b^2}, \sqrt{f^2 - c^2}$$

form the sides of a proper triangle.

A Third Degree Congruence with Nine Solutions

E 2763 [1979, 223]. *Proposed by Lorraine L. Foster, California State University, Northridge.*

Let $f(n) = n^3 + 396n^2 - 111n + 38$. Prove that the congruence $f(n) \equiv 0 \pmod{3^a}$ has precisely 9 solutions $(\operatorname{mod} 3^a)$ for all integers $a \geqslant 5$.

Solution by L. E. Mattics, University of South Alabama.

THEOREM. *If $s \geq 3$, then there is a d such that $f(d) \equiv 0\ (3^{s+2})$ and if $f(b) \equiv 0\ (3^{s+2})$ then $b \equiv d$* $(\operatorname{mod} 3^s)$.

Indeed, $f(n) \equiv 0 \pmod{243}$ if and only if $n \equiv 19 \pmod{27}$ and we proceed by induction assuming the theorem to be true for some $s \geq 3$. Then

$$f(d + d3^s) \equiv f(d) + k3^s f'(d) \pmod{3^{s+3}}$$

and since $d \equiv 1 \pmod 9$, we can show that $3^2 | f'(d)$, but $3^3 \nmid f'(d)$. Hence the congruence

$$f(d) + k3^s f'(d) \equiv 0 \pmod{3^{s+1}}$$

can be solved uniquely for k modulo 3. If $f(c) \equiv 0 \pmod{3^{s+2}}$, then by induction hypothesis $c = a + d \cdot 3^s$ so $d \equiv k \pmod 3$, and $c \equiv a + k3^s \pmod{3^{s+1}}$. The induction is complete and the theorem is proved.

Now if $s \geq 5$ and $f(b) \equiv 0 \pmod{3^s}$, then

$$f(b + m3^{s-2} + r3^{s-1}) \equiv f(b) + m3^{s-2}f'(b) + r3^{s-1}f'(b + m \cdot 3^{s-2}),$$

and this quantity is $\equiv 0 \pmod{3^s}$ because $3^2 | f'(b)$ and $3 | f'(b + m3^{s-2})$. Therefore, since b is unique modulo 3^{s-2}, there are precisely nine solutions modulo 3^s to the congruence $f(n) \equiv 0\ (3^s)$.

Also solved by H. L. Abbott (Canada), J. Ampe (student), L. Kuipers (Switzerland), Jinku Lee (student), N. Miku (Netherlands), A. M. Rockett, R. S. Stacy, P. J. Zwier, and the proposer.

A Product of Powers of an Integer

E 2852 [1980, 672; 1982, 213]. *Proposed by Jan Mycielski, University of Colorado.*

For any positive integer n let $\omega(n)$ be the product of all positive integers k such that n is a power of k. Prove that

$$\prod_{n=1}^{\infty}\left(\frac{\omega(n)}{n+1}\right)^{1/n}=1.$$

Remark by William A. Newcomb, Lawrence Livermore National Laboratory. Neither of the published solutions to this problem [1982, 213] is entirely correct as it stands. They are vitiated, respectively, by the divergence of the series $\Sigma_1^\infty n^{-1}\log(n+1)$, and of the corresponding infinite product, $\Pi_1^\infty(n+1)^{1/n}=\Pi_2^\infty k^{1/(k-1)}$.

A corrected version of Solution I may be given as follows. Let $\omega(n)$ be rewritten as $n\rho(n)$, so that $\rho(n)$ denotes the product of all integers $k<n$ such that n is a power of k. One form of the proposition to be established is

$$\sum_1^\infty \frac{1}{n}\log\rho(n)=\sum_1^\infty\frac{1}{n}\log\frac{n+1}{n}. \tag{1}$$

But now, by essentially the same calculation as in the original Solution I (with ρ replacing ω, and the index a now running from 2 to ∞), one easily obtains

$$\sum_1^\infty\frac{1}{n}\log\rho(n)=\sum_1^\infty\frac{1}{n(n+1)}\log(n+1), \tag{2}$$

from which (1) will follow directly by an application of the Abel transformation. In this version of the argument, all the relevant series are convergent (and indeed, absolutely convergent).

A similar correction may be made to Solution II.

Edmund A. Butler, New Carrollton, Maryland, and the proposer also noticed the omissions. The proposer's solution is as follows.

$$\begin{aligned}\prod_{n=1}^{\infty}\left(\frac{\omega(n)}{n}\right)^{1/n} &= \prod_{\substack{k=2\\ (n\text{ is a power}\\ \text{with exponent}>1)}}^{\infty}\ \prod_{\substack{k^s=n\\ s>1}} k^{1/n}\\ &=\prod_{k=2}^{\infty}\prod_{s=2}^{\infty}k^{1/k^s}\\ &=\prod_{k=2}^{\infty}k^{(1/(k-1))-(1/k)}\\ &=\prod_{n=1}^{\infty}\left(\frac{n+1}{n}\right)^{1/n}.\end{aligned}$$

All the products in this calculation converge absolutely; the last step is also correct because $k^{1/(k-1)}\to 1$. Now we can divide by $\Pi(\frac{n+1}{n})^{1/n}$ and get the desired result.

(This and related identities are stated without proofs in J. Mycielski "Quelques identités de la théorie analytique des nombres," Coll. Math., 4(1956) 68–70.)

Subsets F of Integers for Which $F\subseteq F+F$

E 2899 [1981, 538]. *Proposed by Gerard Letac, Université Paul-Sabatier, Toulouse, France.*

Let n be a positive integer. Consider a set F of distinct integers such that every element of F is

A Product of Powers of an Integer

E 2852 [1980, 672; 1982, 213]. *Proposed by Jan Mycielski, University of Colorado.*

For any positive integer n let $\omega(n)$ be the product of all positive integers k such that n is a power of k. Prove that

$$\prod_{n=1}^{\infty}\left(\frac{\omega(n)}{n+1}\right)^{1/n}=1.$$

Remark by William A. Newcomb, Lawrence Livermore National Laboratory. Neither of the published solutions to this problem [1982, 213] is entirely correct as it stands. They are vitiated, respectively, by the divergence of the series $\sum_1^\infty n^{-1}\log(n+1)$, and of the corresponding infinite product, $\prod_1^\infty(n+1)^{1/n}=\prod_2^\infty k^{1/(k-1)}$.

A corrected version of Solution I may be given as follows. Let $\omega(n)$ be rewritten as $n\rho(n)$, so that $\rho(n)$ denotes the product of all integers $k<n$ such that n is a power of k. One form of the proposition to be established is

$$\sum_1^\infty \frac{1}{n}\log\rho(n)=\sum_1^\infty\frac{1}{n}\log\frac{n+1}{n}. \tag{1}$$

But now, by essentially the same calculation as in the original Solution I (with ρ replacing ω, and the index a now running from 2 to ∞), one easily obtains

$$\sum_1^\infty\frac{1}{n}\log\rho(n)=\sum_1^\infty\frac{1}{n(n+1)}\log(n+1), \tag{2}$$

from which (1) will follow directly by an application of the Abel transformation. In this version of the argument, all the relevant series are convergent (and indeed, absolutely convergent).

A similar correction may be made to Solution II.

Edmund A. Butler, New Carrollton, Maryland, and the proposer also noticed the omissions. The proposer's solution is as follows.

$$\begin{aligned}\prod_{n=1}^{\infty}\left(\frac{\omega(n)}{n}\right)^{1/n} &= \prod_{\substack{k=2\\ (n\text{ is a power}\\ \text{with exponent}>1)}}^{\infty}\ \prod_{\substack{k^s=n\\ s>1}} k^{1/n}\\ &=\prod_{k=2}^{\infty}\prod_{s=2}^{\infty}k^{1/k^s}\\ &=\prod_{k=2}^{\infty}k^{(1/(k-1))-(1/k)}\\ &=\prod_{n=1}^{\infty}\left(\frac{n+1}{n}\right)^{1/n}.\end{aligned}$$

All the products in this calculation converge absolutely; the last step is also correct because $k^{1/(k-1)}\to 1$. Now we can divide by $\prod(\frac{n+1}{n})^{1/n}$ and get the desired result.

(This and related identities are stated without proofs in J. Mycielski "Quelques identités de la théorie analytique des nombres," Coll. Math., 4(1956) 68–70.)

Subsets F of Integers for Which $F\subseteq F+F$

E 2899 [1981, 538]. *Proposed by Gerard Letac, Université Paul-Sabatier, Toulouse, France.*

Let n be a positive integer. Consider a set F of distinct integers such that every element of F is

the sum of two elements of F and the sum of k elements of F is never 0 for $k = 1, 2, \ldots, n$. Prove that F must have at least $2n + 2$ elements. (*) Find all F with $|F| = 2n + 2$.

Solution by Allen J. Schwenk, University of Waterloo. Partition F into a positive and negative subset, $F = F^+ \cup F^-$. Define a directed graph on F^+ by inserting an arc from x to y if $y = x + z$ for some $z \in F$. (Note that x and y are positive but z may be positive or negative.) By hypothesis, every vertex has positive indegree. Such a graph must contain a directed cycle, say $p_1 p_2 \cdots p_k p_1$. Now we have

$$\begin{aligned} p_1 &= p_k + z_k \\ &= p_{k-1} + z_{k-1} + z_k \\ &= p_{k-2} + z_{k-2} + z_{k-1} + z_k \\ &\;\;\vdots \\ &= p_1 + z_1 + z_2 + \cdots + z_k. \end{aligned}$$

Subtracting p_1 yields 0 as a sum of k elements, so $k \geq n + 1$. Thus $|F^+| \geq n + 1$. A similar argument applies to F^-.

Now if $|F| = 2n + 2$, the directed graphs on F^+ and F^- must each be precisely a directed $n + 1$ cycle, since any additional arc would create a shorter directed cycle and hence a shorter 0 sum. Consequently, each element in the set is realized as a sum in a unique way, and if $y = x + z$ with x, y, and z all having the same sign, then $x = z$, for otherwise y has indegree 2 violating the structure of F^+ or F^-. Observe that if F is any set with the prescribed properties, then so is mF for any nonzero multiplier m.

To construct such sets for each positive n, first select p_0 to be a positive integer such that $p_0 < 2^{n+1} - 1$ and $\gcd(p_0, 2^{n+1} - 1) = 1$. Define p_k by

$$p_k \equiv 2^k p_0 \bmod (2^{n+1} - 1) \quad \text{and} \quad 0 < p_k < 2^{n+1} - 1.$$

Let $F^+ = \{p_0, p_1, \ldots, p_n\}$ and set $F^- = \{q_k | q_k = p_k + 1 - 2^{n+1}\}$. To see that each element can be realized as a sum, observe that if $p_k < 2^n$, then $p_{k+1} = 2p_k$ and

$$\begin{aligned} q_{k+1} &= p_{k+1} - (2^{n+1} - 1) \\ &= 2p_k - (2^{n+1} - 1) \\ &= q_k + p_k. \end{aligned}$$

On the other hand, if $p_k \geq 2^n$, then

$$\begin{aligned} p_{k+1} &= 2p_k - (2^{n+1} - 1) = p_k + q_k \quad \text{and} \\ q_{k+1} &= p_{k+1} - (2^{n+1} - 1) \\ &= 2p_k - 2(2^{n+1} - 1) \\ &= 2q_k. \end{aligned}$$

To show that there are no forbidden zero sums, suppose $x_1 + x_2 + \cdots + x_k = 0$ with each $x_i \in F$. Read this equation modulo $2^{n+1} - 1$. After multiplying by p_0^{-1}, each term is a power 2^i with $0 \leq i \leq n$. Clearly at least $n + 1$ terms must be present to produce 0 mod $2^{n+1} - 1$. It remains to decide whether the sets just constructed and their multiples are the only possible sets of minimum order $2n + 2$.

Let $F_n = F_n^+ \cup F_n^-$ denote a set with the prescribed properties and order $2n + 2$. By scaling with a suitable multiplier, we may presume that the element of smallest magnitude is $+1$. Let $p_0 = 1$. Since $p_1 = p_0 + z_0$, we cannot have $z_0 < 0$, for then $|z_0| < 1$, so $z_0 = 1$ and p_1 must be 2.

Now if $n = 1$, $p_0 = p_1 + z_1$ implies $-1 \in F$, so $q_0 = -1$. Then $q_1 = q_0 + w_0$ implies $q_1 = -2$. Thus $F_2 = m\{1, 2, -1, -2\}$ is essentially unique.

For $n \geq 2$, we may continue with $p_2 = p_1 + z_1$. If $z_1 < 0$, then $|z_1| < 1$ is a contradiction. Thus $z_1 = p_1$ and $p_2 = 4$. If $n = 2$, $p_0 = p_2 + z_2$ implies $-3 \in F^-$. Since F^+ and one element of F^- are all integral, the cyclic generation of the rest of F^- assures that it too is integral. (Our earlier rescaling could have destroyed the hypothesized integrality.) Since -1 and -2 would produce two term 0 sums, -3 is the element of least magnitude in F^-. Thus $q_0 = -3$ implies that $q_1 = 2q_0 = -6$ and then $q_2 = q_1 + w_1$. If $w_1 = q_1$ we cannot cycle back to $q_0 = -3$ in one more step. Thus $w_1 \in F_2^+$. The only choice that avoids contradiction is $q_2 = -5$. Thus $F_2 = m\{1, 2, 4, -3, -6, -5\}$ is unique.

For $n = 3$, we have $p_3 = p_2 + z_2$. If $z_2 < 0$, then $-1 < z_2 < -2$ since $p_0 = p_3 + z_3$ with $z_3 < 0$. Now $z_2 = z_3 = -1.5$ implies that $q_0 = -1.5$ and so $q_1 = -3$ yielding the sum $0 = 1 + 2 - 3$. Thus, z_2 and z_3 are distinct elements of F_3^- which sum to -3. Consequently, $|z_2 - z_3| < 1$ which requires that these elements not be consecutive on the cycle in F_3^-. Thus $\{z_2, z_3\} = \{q_0, q_2\}$. Now q_1 must be $2q_0$ and $q_2 = q_1 + w_1$ with $w_1 \in F_3^+$. The only legal choice is $w_1 = 1$ so $q_2 = 2q_0 + 1$ with $-1 < q_2 < -2$. Next $q_3 = q_2 + w_2$ cannot tolerate a positive w_2, so $q_3 = 2q_2 = 4q_0 + 2$. Finally, $q_0 = q_3 + w_3$ requires $-3q_0 - 2 \in F_3^+$. Selecting $-3q_0 - 2 = 1$ or 4 yields q_0 outside the required range $-1 < q_0 < -1.5$. If we select $-3q_0 - 2 = 2$ we find $F_3^- = \{-\frac{4}{3}, -\frac{8}{3}, -\frac{5}{3}, -\frac{10}{3}\}$. But now $p_3 = \frac{7}{3}$ or $\frac{8}{3}$ with each choice producing a forbidden 0 sum. Thus, $-3q_0 - 2 = p_3 = 4 + z_2$. Either choice $z_2 = q_0$ or q_2 yields $q_0 < -2$, so the initial assumption $z_2 < 0$ cannot occur. Thus F_3^+ must be $\{1, 2, 4, 8\}$ and $-7 \in F_3^-$. As with F_2, we observe that F_3^- must be integral with $q_0 = -7$ being the element of smallest magnitude. Then $q_1 = -14$, $q_2 = -13$, and $q_3 = -11$ are all forced. Consequently, $F_3 = m\{1, 2, 4, 8, -7, -14, -13, -11\}$ is unique.

For each $n \geq 4$ there will be more than one possible set. For example, with $n = 4$ we have already constructed:

$$F_4 = m\{1, 2, 4, 8, 16, -30, -29, -27, -23, -15\},$$
$$F_4 = m\{3, 6, 12, 24, 17, -28, -25, -19, -7, -14\}, \quad \text{and}$$
$$F_4 = m\{5, 10, 20, 9, 18, -26, -21, -11, -22, -13\}.$$

It is tempting to conjecture that these will be the only possible sets. For $n \geq 4$ this question remains open.

Additional acknowledgements

E 2747. Also solved by A. A. Jagers (Netherlands).
E 2762. Also solved by F. Gerrish (England).
E 2780. Also solved by M. W. Ecker, J. E. Morrill.
E 2781. Also solved by R. Breusch, D. M. Broline, M. W. Ecker, M. Gerstell, E. Hensley, R. Hill (United Kingdom), M. S. Jacobson, S. B. Maurer, E. Trost (Switzerland, G. Shulman, S. Wagon.
E 2782. Also solved by L. Kuipers (Switzerland), R. Lyons.
E 2783. Also solved by K. S. Alexander, F. S. Cater, M. W. Ecker, D. Hensley, B. Klein, J. D. McCall, S. V. Noltie, N. D. Ritter, J. H. Silverman, J. C. Smith, J. Wiener & P. Pontius, P. Zwier.
E 2792. Also solved by Robert Patenaude.
E 2806. Also solved by N. E. Frangos, V. Pambuccian (Rumania).
E 2835. Also solved by H. N. Gupta, R. J. Tomkins, P. Zwier.
E 2845. Also solved by D. K. Ross (Australia).
E 2851. Also solved by S. A. Heuer, D. B. Shapiro.
E 2857. Also solved by Elgin Johnston.
E 2859. Also solved by R. Bournas.
E 2866. Also solved by S. A. Belbas, K. L. Bernstein, J. Binz (Switzerland), J. N. Boyd, N. Brown (student, Australia), P. S. Bruckman, M. P. Eisner, H. Eves, H. E. Fettis, C. Gengzhe, P. Hajnal (Hungary), H. Kappus (Germany), E. L. Koh (Saudi Arabia), D. Logothetti, O. P. Lossers (Netherlands), H. J. Ludwig, H. M. Marston, N. Martin (Canada), W. W. Meyer, M. W. Meyerson, W. Moser (Canada), W. A. Newcomb, J. Papès-F. (Spain), J. P. Robertson, I. A. Sakmar (Canada), B. Shelupsky, N. S. Shou (Hong Kong), J. Suck (Germany), H. R. van der Vaart, an anonymous solver, and the proposer.

E 2950. Also solved by N. Franceschine, III, O. P. Lossers (Netherlands); L. E. Mattics, N. Miku (Netherlands), D. E. Orr, and the proposer.

6274. Also solved by V. Pambuccian (Rumania).

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by May 31, 1983. The solver's full post-office address should be on each sheet.

6415. *Proposed by Raphael M. Robinson, University of California, Berkeley.*

If $P(x_1, x_2, \ldots, x_n)$ is a polynomial with integer coefficients, find m and a homogeneous polynomial $Q(x_1, x_2, \ldots, x_m)$ with integer coefficients, so that $P = 0$ has a solution in rationals if and only if $Q = 0$ has a solution in integers not all zero.

6416. *Proposed by Sanford S. Miller, SUNY, Brockport, NY.*

Let $p(z) = 1 + p_1 z + \cdots$ be regular in the unit disc U, and let α be a real number such that $\alpha < 1$. Show that if

$$\operatorname{Re}\left[z^2 p''(z) + 3zp'(z) - \alpha p^2(z)\right] > -1$$

for $z \in U$, then $\operatorname{Re} p(z) > 0$.

6417. *Proposed by R. P. Boas, Northwestern University.*

It is a well-known theorem of Pólya that although the cosine transform of a decreasing positive function is not necessarily positive, the cosine transform of a positive convex function is positive. [Math. Z., 2 (1918) 352-382 (378) = Collected Papers, vol. 2, M.I.T. Press, 1974, p. 192.] Show that there is a positive number $0 < \sigma < \frac{1}{2}$ such that the cosine transform of a positive function ϕ with $t^{\mu}\phi(t)$ decreasing is positive provided that $\mu > \sigma$ but not necessarily when $\mu < \sigma$. (Note that, although $\phi(t)$ necessarily becomes infinite at $t = 0$, $\phi(t)$ might be 0 for $t > t_0$.)

6418. *Proposed by George Benke, Georgetown University.*

Prove that

$$\sum_{n=1}^{2N-1} \frac{\operatorname{Sin} \dfrac{\pi n^2}{2N}}{\operatorname{Sin} \dfrac{\pi n}{2N}} = N.$$

6419. *Proposed by Wim Vervaat, Cornell University.*

Let f be a real-valued Lebesgue measurable function on $\mathbb{R}$.

(a) Suppose that $f(qx) = qf(x)$ for all $q \in \mathbb{Z}$, $x \in \mathbb{R}$. Prove that $f(x) = bx$ almost everywhere on $\mathbb{R}$ for some real b.

(b) If $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$, then there is a real b such that $f(x) = bx$ for all $x \in \mathbb{R}$. Many proofs are known for this result. Prove it this time as a corollary of (a).

6420. *Proposed by Barry Powell, Kirkland, WA.*

For any odd prime p, let g_p and G_p be the least positive and greatest negative primitive root of p respectively. Prove that for any positive integer M, there exist infinitely many primes p for which $M < g_p < (p - M)$ and $(-p + M) < G_p < -M$.

E 2950. Also solved by N. Franceschine, III, O. P. Lossers (Netherlands); L. E. Mattics, N. Miku (Netherlands), D. E. Orr, and the proposer.

6274. Also solved by V. Pambuccian (Rumania).

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by May 31, 1983. The solver's full post-office address should be on each sheet.

6415. *Proposed by Raphael M. Robinson, University of California, Berkeley.*

If $P(x_1, x_2, \ldots, x_n)$ is a polynomial with integer coefficients, find m and a homogeneous polynomial $Q(x_1, x_2, \ldots, x_m)$ with integer coefficients, so that $P = 0$ has a solution in rationals if and only if $Q = 0$ has a solution in integers not all zero.

6416. *Proposed by Sanford S. Miller, SUNY, Brockport, NY.*

Let $p(z) = 1 + p_1 z + \cdots$ be regular in the unit disc U, and let α be a real number such that $\alpha < 1$. Show that if

$$\operatorname{Re}\left[z^2 p''(z) + 3zp'(z) - \alpha p^2(z)\right] > -1$$

for $z \in U$, then $\operatorname{Re} p(z) > 0$.

6417. *Proposed by R. P. Boas, Northwestern University.*

It is a well-known theorem of Pólya that although the cosine transform of a decreasing positive function is not necessarily positive, the cosine transform of a positive convex function is positive. [Math. Z., 2 (1918) 352-382 (378) = Collected Papers, vol. 2, M.I.T. Press, 1974, p. 192.] Show that there is a positive number $0 < \sigma < \frac{1}{2}$ such that the cosine transform of a positive function ϕ with $t^\mu \phi(t)$ decreasing is positive provided that $\mu > \sigma$ but not necessarily when $\mu < \sigma$. (Note that, although $\phi(t)$ necessarily becomes infinite at $t = 0$, $\phi(t)$ might be 0 for $t > t_0$.)

6418. *Proposed by George Benke, Georgetown University.*

Prove that

$$\sum_{n=1}^{2N-1} \frac{\operatorname{Sin} \dfrac{\pi n^2}{2N}}{\operatorname{Sin} \dfrac{\pi n}{2N}} = N.$$

6419. *Proposed by Wim Vervaat, Cornell University.*

Let f be a real-valued Lebesgue measurable function on $\mathbb{R}$.

(a) Suppose that $f(qx) = qf(x)$ for all $q \in \mathbb{Z}$, $x \in \mathbb{R}$. Prove that $f(x) = bx$ almost everywhere on $\mathbb{R}$ for some real b.

(b) If $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$, then there is a real b such that $f(x) = bx$ for all $x \in \mathbb{R}$. Many proofs are known for this result. Prove it this time as a corollary of (a).

6420. *Proposed by Barry Powell, Kirkland, WA.*

For any odd prime p, let g_p and G_p be the least positive and greatest negative primitive root of p respectively. Prove that for any positive integer M, there exist infinitely many primes p for which $M < g_p < (p - M)$ and $(-p + M) < G_p < -M$.

SOLUTIONS OF ADVANCED PROBLEMS

Collinear Algebraic Units

6341 [1981, 294]. *Proposed by J. C. Lagarias, Bell Laboratories, and H. W. Lenstra, University of Amsterdam.*

Let L be a line in the complex plane C passing through two relatively prime algebraic integers. (Two algebraic integers α, β are *relative prime* if there exist algebraic integers γ, δ such that $\alpha\delta + \beta\gamma = 1$.) Prove that there are infinitely many algebraic units lying on the line L.

Solution by David G. Cantor, University of California, Los Angeles. Suppose α, β are relatively prime algebraic integers. It is enough to prove there exist infinitely many $t \in \mathbb{R}$ such that $\alpha + \beta t$ is a unit. Let K be the (real) field obtained by adjoining the real and imaginary parts of α and β to $\mathbb{Q}$. Let L be a real finite extension of K containing at least two Archimedean valuations. Let v_1 be the valuation of L obtained by the given embedding of L into $\mathbb{R}$. Let v_2 be another Archimedean valuation of L. It suffices to prove there exist infinitely many algebraic $t \in \mathbb{R}$ such that $|\alpha + \beta t|_v = 1$ for all non-Archimedean valuations. Since no restriction is placed on the value of t under v_2, we can even specify that t lie in an arbitrary small interval (r, s). The result is now an extremely special case of Theorem 5.1.1 of D. G. Cantor, *On an extension of the definition of transfinite diameter and some applications*, J. für die reine und angewandte Mathematik, 316(1980) 160–207.

Also solved by the proposers.

Units of Chains

6342 [1981, 294]. *Proposed by Richard Stanley, Massachusetts Institute of Technology.*

Let $f(n)$ be the number of nonisomorphic n-element partially ordered sets P which do not contain three pairwise incomparable elements. (Equivalently, P is a union of two chains.) Let

$$F(x) = 1 + \sum_{n \geqslant 1} f(n)x^n = 1 + x + 2x^2 + 4x^3 + 10x^4 + \cdots .$$

Show that

$$F(x) = \frac{4}{2 - 2x + \sqrt{1 - 4x} + \sqrt{1 - 4x^2}}.$$

Solution by the proposer. Let $g(n)$ be the number of such P that do not contain a proper subset Q such that $x \leqslant y$ for all $x \in Q$ and $y \notin Q$. Let

$$G(x) = \sum_{n \geqslant 1} g(n)x^n = x + x^2 + x^3 + 3x^4 + \cdots .$$

It is easily seen that

$$F(x) = (1 - G(x))^{-1}.$$

A poset (= partially ordered set) P enumerated by $g(n)$ can be written uniquely (apart from order) as the disjoint union of two chains C_1 and C_2. Let C_1 be given by $x_1 < x_2 < \cdots < x_r$ and C_2 by $y_1 < y_2 < \cdots < y_s$. Define $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ by

$$a_i = \begin{cases} 1, & \text{if for some } x_j,\quad i = |\{x \in P | x \leqslant x_j\}|, \\ -1, & \text{otherwise,} \end{cases}$$

SOLUTIONS OF ADVANCED PROBLEMS

Collinear Algebraic Units

6341 [1981, 294]. *Proposed by J. C. Lagarias, Bell Laboratories, and H. W. Lenstra, University of Amsterdam.*

Let L be a line in the complex plane C passing through two relatively prime algebraic integers. (Two algebraic integers α, β are *relative prime* if there exist algebraic integers γ, δ such that $\alpha\delta + \beta\gamma = 1$.) Prove that there are infinitely many algebraic units lying on the line L.

Solution by David G. Cantor, University of California, Los Angeles. Suppose α, β are relatively prime algebraic integers. It is enough to prove there exist infinitely many $t \in \mathbb{R}$ such that $\alpha + \beta t$ is a unit. Let K be the (real) field obtained by adjoining the real and imaginary parts of α and β to $\mathbb{Q}$. Let L be a real finite extension of K containing at least two Archimedean valuations. Let v_1 be the valuation of L obtained by the given embedding of L into $\mathbb{R}$. Let v_2 be another Archimedean valuation of L. It suffices to prove there exist infinitely many algebraic $t \in \mathbb{R}$ such that $|\alpha + \beta t|_v = 1$ for all non-Archimedean valuations. Since no restriction is placed on the value of t under v_2, we can even specify that t lie in an arbitrary small interval (r, s). The result is now an extremely special case of Theorem 5.1.1 of D. G. Cantor, *On an extension of the definition of transfinite diameter and some applications*, J. für die reine und angewandte Mathematik, 316(1980) 160–207.

Also solved by the proposers.

Units of Chains

6342 [1981, 294]. *Proposed by Richard Stanley, Massachusetts Institute of Technology.*

Let $f(n)$ be the number of nonisomorphic n-element partially ordered sets P which do not contain three pairwise incomparable elements. (Equivalently, P is a union of two chains.) Let

$$F(x) = 1 + \sum_{n \geqslant 1} f(n)x^n = 1 + x + 2x^2 + 4x^3 + 10x^4 + \cdots .$$

Show that

$$F(x) = \frac{4}{2 - 2x + \sqrt{1-4x} + \sqrt{1-4x^2}}.$$

Solution by the proposer. Let $g(n)$ be the number of such P that do not contain a proper subset Q such that $x \leqslant y$ for all $x \in Q$ and $y \notin Q$. Let

$$G(x) = \sum_{n \geqslant 1} g(n)x^n = x + x^2 + x^3 + 3x^4 + \cdots .$$

It is easily seen that

$$F(x) = (1 - G(x))^{-1}.$$

A poset (= partially ordered set) P enumerated by $g(n)$ can be written uniquely (apart from order) as the disjoint union of two chains C_1 and C_2. Let C_1 be given by $x_1 < x_2 < \cdots < x_r$ and C_2 by $y_1 < y_2 < \cdots < y_s$. Define $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ by

$$a_i = \begin{cases} 1, & \text{if for some } x_j,\ \ i = |\{x \in P \mid x \leqslant x_j\}|, \\ -1, & \text{otherwise,} \end{cases}$$

$$b_i = \begin{cases} -1, & \text{if for some } y_j, \quad i = |\{y \in P | y \leqslant y_j\}|, \\ 1, & \text{otherwise.} \end{cases}$$

Then $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ satisfy $a_1 = 1, b_1 = -1,$

$$a_1 + \cdots + a_i \neq b_1 + \cdots + b_i \text{ for } 1 \leqslant i < n,$$

and

$$a_1 + \cdots + a_n = b_1 + \cdots + b_n.$$

The number of such pairs of sequences is known to be the Catalan number

$$C_{n-1} = \frac{1}{n}\binom{2(n-1)}{n-1}.$$

Each such pair yields a poset P enumerated by $g(n)$, and two sequences yield the same poset P if and only if $a_i = -b_i$ for all i. The number of sequences $a_1, a_2, \ldots, a_n$ such that $a_1 = 1$,

$$a_1 + \cdots + a_i \neq -(a_1 + \cdots + a_i) \text{ for } 1 \leqslant i < n$$

and

$$a_1 + \cdots + a_n = -(a_1 + \cdots + a_n)$$

is 0 if n is odd and is $C_{n/2-1}$ if n is even. Hence

$$g(2n) = \tfrac{1}{2}(C_{2n-1} + C_{n-1})$$
$$g(2n+1) = \tfrac{1}{2}C_{2n}.$$

The proof now follows straightforwardly from the well-known generating function

$$\sum_{n \geqslant 0} C_n x^n = \frac{1 - \sqrt{1-4x}}{2x}.$$

A Translation Invariant Measure

6348 [1981, 446]. *Proposed by James Henle and Stanley Wagon, Smith College.*

Prove or disprove: If μ is a countably additive translation invariant measure from the Borel subsets of $\mathbb{R}$ to $[0, \infty]$, then μ is invariant under all isometries of $\mathbb{R}$; i.e., it is also invariant under reflections.

Solution by the proposers. The assertion is false, although it is true under the additional hypothesis $\mu([0,1]) < \infty$, for then μ is a scalar multiple of Lebesgue measure. To get a counterexample it suffices to construct a Borel set $C \subseteq [0,1]$ such that for any countable set, $\{a_n\}$, of reals,

$$-C \nsubseteq \cup \{C + a_n : n < \infty\}.$$

For then we may define a σ-ideal I of Borel sets by:

$$I = \left\{A \subseteq \mathbb{R} : A \text{ is Borel and } A \subseteq \bigcup_{n=1}^{\infty} C + a_n \text{ for some sequence of reals } a_n\right\}.$$

Then let μ be the measure assigning measure 0 to all sets in I, and measure ∞ to all other Borel sets. This is the desired measure since μ is translation invariant and countably additive, $\mu(C) = 0$, and $\mu(-C) = \infty$.

It remains to construct C. Simply let C consist of those reals in $[0,1]$ having a base 4 expansion using no 2's. Then C is an asymmetric Cantor set, which is closed, and hence Borel. To prove that C has the desired property it suffices to show that, for any sequence $\{a_n\}$,

$$b_i = \begin{cases} -1, & \text{if for some } y_j, \quad i = |\{y \in P | y \leqslant y_j\}|, \\ 1, & \text{otherwise.} \end{cases}$$

Then $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ satisfy $a_1 = 1, b_1 = -1$,

$$a_1 + \cdots + a_i \neq b_1 + \cdots + b_i \text{ for } 1 \leqslant i < n,$$

and

$$a_1 + \cdots + a_n = b_1 + \cdots + b_n.$$

The number of such pairs of sequences is known to be the Catalan number

$$C_{n-1} = \frac{1}{n}\binom{2(n-1)}{n-1}.$$

Each such pair yields a poset P enumerated by $g(n)$, and two sequences yield the same poset P if and only if $a_i = -b_i$ for all i. The number of sequences $a_1, a_2, \ldots, a_n$ such that $a_1 = 1$,

$$a_1 + \cdots + a_i \neq -(a_1 + \cdots + a_i) \text{ for } 1 \leqslant i < n$$

and

$$a_1 + \cdots + a_n = -(a_1 + \cdots + a_n)$$

is 0 if n is odd and is $C_{n/2-1}$ if n is even. Hence

$$g(2n) = \tfrac{1}{2}(C_{2n-1} + C_{n-1})$$
$$g(2n+1) = \tfrac{1}{2}C_{2n}.$$

The proof now follows straightforwardly from the well-known generating function

$$\sum_{n \geqslant 0} C_n x^n = \frac{1 - \sqrt{1-4x}}{2x}.$$

A Translation Invariant Measure

6348 [1981, 446]. *Proposed by James Henle and Stanley Wagon, Smith College.*

Prove or disprove: If μ is a countably additive translation invariant measure from the Borel subsets of $\mathbb{R}$ to $[0, \infty]$, then μ is invariant under all isometries of $\mathbb{R}$; i.e., it is also invariant under reflections.

Solution by the proposers. The assertion is false, although it is true under the additional hypothesis $\mu([0,1]) < \infty$, for then μ is a scalar multiple of Lebesgue measure. To get a counterexample it suffices to construct a Borel set $C \subseteq [0,1]$ such that for any countable set, $\{a_n\}$, of reals,

$$-C \nsubseteq \cup \{C + a_n : n < \infty\}.$$

For then we may define a σ-ideal I of Borel sets by:

$$I = \left\{A \subseteq \mathbb{R} : A \text{ is Borel and } A \subseteq \bigcup_{n=1}^{\infty} C + a_n \text{ for some sequence of reals } a_n\right\}.$$

Then let μ be the measure assigning measure 0 to all sets in I, and measure ∞ to all other Borel sets. This is the desired measure since μ is translation invariant and countably additive, $\mu(C) = 0$, and $\mu(-C) = \infty$.

It remains to construct C. Simply let C consist of those reals in $[0,1]$ having a base 4 expansion using no 2's. Then C is an asymmetric Cantor set, which is closed, and hence Borel. To prove that C has the desired property it suffices to show that, for any sequence $\{a_n\}$,

$$1 - C \not\subseteq \bigcup_{n=1}^{\infty} C + a_n.$$

Note that $1 - C$ consists of reals in [0, 1] having a base 4 expansion using no 1's. Suppose $\{a_n\}$ is given. It is a simple matter to get a real x in [0, 1] by building its base 4 expansion inductively so that the expansion has no 1's (and hence x is $1 - C$) but yet, for each n, $x - a_n$ has two 2's in a base 4 expansion, and therefore at least one 2 in any of its base 4 expansions. Namely, suppose the first r base 4 digits of x have been constructed guaranteeing that x is as required with respect to a_m for $m < n$, and let x_{n-1} denote the number beginning with the already constructed r digits of x and ending in a string of 0's.

Consider the base 4 expansion (not ending in all 3's) of $x_{n-1} - a_n$. If this expansion ends in all 1's, add a string of the form $0\ldots 03030$ to the already constructed digits of x, where the string 3030 corresponds to positions beyond the rth where $x_{n-1} - a_n$ has 1's. Otherwise add a string of the form $0\ldots 0i0\ldots 0\ldots 0j0\ldots 0$ to x where i, j are in positions beyond the rth where $x_{n-1} - a_n$ has a 0, 2 or 3, and i, j equal 2, 0 or 3 accordingly; moreover, the final 0 and at least one of the 0's between i and j are in positions where $x_{n-1} - a_n$ has a digit different from 3 (this is so that carrying will not enter into the arithmetic). These additions guarantee that $x - a_n$ will have a 2 in any of its base 4 expansions, and hence that x will not lie in $C + a_n$.

K. E. Hirst points out that a counterexample is given in his paper "Translation invariant measures which are not Hausdorff measures," Proc. Cambridge Philos. Soc., 62 (1966) 693–698.

Countable Sets of Linear Functions

6349 [1981, 446]. *Proposed by J.-B. Hiriart-Urruty, University of Clermont-Ferrand II, France, and T. H. Sweetser, University of California at San Diego.*

Let $\mathcal{A}$ be a countable set in $L(\mathbb{R}^n, \mathbb{R}^m)$—the space of all linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$; we define $\mathcal{A}x = \{Ax \mid A \in \mathcal{A}\}$. Does $B \in L(\mathbb{R}^n, \mathbb{R}^m)$ satisfying $Bx \in \mathcal{A}x \forall x \in \mathbb{R}^n$ necessarily belong to $\mathcal{A}$?

Solution by Chin-Chi Shan, Ohio State University. The answer is "yes." Let $\{e_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$ and

$$W = \left\{ \sum_{i=1}^{n} a^{i-1} e_i : a \in \mathbb{R} \setminus \{0\} \right\}.$$

Then any n vectors in W constitute a basis of $\mathbb{R}^n$ since their coefficients from a Vandermonde determinant. Since W is uncountable and $\mathcal{A}$ is countable, there exists at least one $A \in \mathcal{A}$ such that $|\{x \in W: Bx = Ax\}| \geqslant n$. Hence $B = A$ on a basis of $\mathbb{R}^n$, and so $B = A$ on $\mathbb{R}^n$.

Also solved by V. Anantharam, Carlos Biasi and Wilson Mauricio Tadini (Brazil), Jozef Bobok (Czechoslovakia), Jerzy Browkin (Poland), F. S. Cater, Chico Problem Group, John Cobb, G. Crofts, Patrick R. Gardner, Jaroslav Guričan (Czechoslovakia), William H. Gustafson, Dinh Thê Hung and Benny Cheng, Elgin Johnston, Joel Levy, J. G. Mauldon, William Myers, Ivan Netuka (Czechoslovakia), Nicholas Passell, Eero Posti (Finland), Adam Riese, C. Ray Rosentrater, A. R. Sourour (Canada), R. A. Vitale, David M. Wells, J. B. Wilker (Canada), Pei Yuan Wu (Taiwan), and the proposers.

Separation of Edges of a Square by a Closed Set

6351 [1981, 540]. *Proposed by James W. Walker, MIT.*

Let F be a closed subset of the unit square $I^2 = [0,1] \times [0,1]$ such that F contains the closed horizontal edges $[0,1] \times \{0\}$ and $[0,1] \times \{1\}$, but does not intersect the open vertical edges $\{0\} \times (0,1)$ and $\{1\} \times (0,1)$. Show that either both horizontal edges belong to the same connected

$$1 - C \not\subseteq \bigcup_{n=1}^{\infty} C + a_n.$$

Note that $1 - C$ consists of reals in $[0, 1]$ having a base 4 expansion using no 1's. Suppose $\{a_n\}$ is given. It is a simple matter to get a real x in $[0, 1]$ by building its base 4 expansion inductively so that the expansion has no 1's (and hence x is $1 - C$) but yet, for each n, $x - a_n$ has two 2's in a base 4 expansion, and therefore at least one 2 in any of its base 4 expansions. Namely, suppose the first r base 4 digits of x have been constructed guaranteeing that x is as required with respect to a_m for $m < n$, and let x_{n-1} denote the number beginning with the already constructed r digits of x and ending in a string of 0's.

Consider the base 4 expansion (not ending in all 3's) of $x_{n-1} - a_n$. If this expansion ends in all 1's, add a string of the form $0\ldots 03030$ to the already constructed digits of x, where the string 3030 corresponds to positions beyond the rth where $x_{n-1} - a_n$ has 1's. Otherwise add a string of the form $0\ldots 0i0\ldots 0\ldots 0j0\ldots 0$ to x where i, j are in positions beyond the rth where $x_{n-1} - a_n$ has a 0, 2 or 3, and i, j equal 2, 0 or 3 accordingly; moreover, the final 0 and at least one of the 0's between i and j are in positions where $x_{n-1} - a_n$ has a digit different from 3 (this is so that carrying will not enter into the arithmetic). These additions guarantee that $x - a_n$ will have a 2 in any of its base 4 expansions, and hence that x will not lie in $C + a_n$.

K. E. Hirst points out that a counterexample is given in his paper "Translation invariant measures which are not Hausdorff measures," Proc. Cambridge Philos. Soc., 62 (1966) 693–698.

Countable Sets of Linear Functions

6349 [1981, 446]. *Proposed by J.-B. Hiriart-Urruty, University of Clermont-Ferrand II, France, and T. H. Sweetser, University of California at San Diego.*

Let $\mathcal{A}$ be a countable set in $L(\mathbb{R}^n, \mathbb{R}^m)$—the space of all linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$; we define $\mathcal{A}x = \{Ax | A \in \mathcal{A}\}$. Does $B \in L(\mathbb{R}^n, \mathbb{R}^m)$ satisfying $Bx \in \mathcal{A}x \forall x \in \mathbb{R}^n$ necessarily belong to $\mathcal{A}$?

Solution by Chin-Chi Shan, Ohio State University. The answer is "yes." Let $\{e_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$ and

$$W = \left\{ \sum_{i=1}^{n} a^{i-1} e_i : a \in \mathbb{R} \setminus \{0\} \right\}.$$

Then any n vectors in W constitute a basis of $\mathbb{R}^n$ since their coefficients from a Vandermonde determinant. Since W is uncountable and $\mathcal{A}$ is countable, there exists at least one $A \in \mathcal{A}$ such that $|\{x \in W: Bx = Ax\}| \geqslant n$. Hence $B = A$ on a basis of $\mathbb{R}^n$, and so $B = A$ on $\mathbb{R}^n$.

Also solved by V. Anantharam, Carlos Biasi and Wilson Mauricio Tadini (Brazil), Jozef Bobok (Czechoslovakia), Jerzy Browkin (Poland), F. S. Cater, Chico Problem Group, John Cobb, G. Crofts, Patrick R. Gardner, Jaroslav Guričan (Czechoslovakia), William H. Gustafson, Dinh Thế Hưng and Benny Cheng, Elgin Johnston, Joel Levy, J. G. Mauldon, William Myers, Ivan Netuka (Czechoslovakia), Nicholas Passell, Eero Posti (Finland), Adam Riese, C. Ray Rosentrater, A. R. Sourour (Canada), R. A. Vitale, David M. Wells, J. B. Wilker (Canada), Pei Yuan Wu (Taiwan), and the proposers.

Separation of Edges of a Square by a Closed Set

6351 [1981, 540]. *Proposed by James W. Walker, MIT.*

Let F be a closed subset of the unit square $I^2 = [0, 1] \times [0, 1]$ such that F contains the closed horizontal edges $[0, 1] \times \{0\}$ and $[0, 1] \times \{1\}$, but does not intersect the open vertical edges $\{0\} \times (0, 1)$ and $\{1\} \times (0, 1)$. Show that either both horizontal edges belong to the same connected

$$1 - C \not\subseteq \bigcup_{n=1}^{\infty} C + a_n.$$

Note that $1 - C$ consists of reals in $[0, 1]$ having a base 4 expansion using no 1's. Suppose $\{a_n\}$ is given. It is a simple matter to get a real x in $[0, 1]$ by building its base 4 expansion inductively so that the expansion has no 1's (and hence x is $1 - C$) but yet, for each n, $x - a_n$ has two 2's in a base 4 expansion, and therefore at least one 2 in any of its base 4 expansions. Namely, suppose the first r base 4 digits of x have been constructed guaranteeing that x is as required with respect to a_m for $m < n$, and let x_{n-1} denote the number beginning with the already constructed r digits of x and ending in a string of 0's.

Consider the base 4 expansion (not ending in all 3's) of $x_{n-1} - a_n$. If this expansion ends in all 1's, add a string of the form $0\ldots 03030$ to the already constructed digits of x, where the string 3030 corresponds to positions beyond the rth where $x_{n-1} - a_n$ has 1's. Otherwise add a string of the form $0\ldots 0i0\ldots 0\ldots 0j0\ldots 0$ to x where i, j are in positions beyond the rth where $x_{n-1} - a_n$ has a 0, 2 or 3, and i, j equal 2, 0 or 3 accordingly; moreover, the final 0 and at least one of the 0's between i and j are in positions where $x_{n-1} - a_n$ has a digit different from 3 (this is so that carrying will not enter into the arithmetic). These additions guarantee that $x - a_n$ will have a 2 in any of its base 4 expansions, and hence that x will not lie in $C + a_n$.

K. E. Hirst points out that a counterexample is given in his paper "Translation invariant measures which are not Hausdorff measures," Proc. Cambridge Philos. Soc., 62 (1966) 693–698.

Countable Sets of Linear Functions

6349 [1981, 446]. *Proposed by J.-B. Hiriart-Urruty, University of Clermont-Ferrand II, France, and T. H. Sweetser, University of California at San Diego.*

Let $\mathcal{Q}$ be a countable set in $L(\mathbb{R}^n, \mathbb{R}^m)$—the space of all linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$; we define $\mathcal{Q}x = \{Ax | A \in \mathcal{Q}\}$. Does $B \in L(\mathbb{R}^n, \mathbb{R}^m)$ satisfying $Bx \in \mathcal{Q}x \forall x \in \mathbb{R}^n$ necessarily belong to $\mathcal{Q}$?

Solution by Chin-Chi Shan, Ohio State University. The answer is "yes." Let $\{e_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$ and

$$W = \left\{ \sum_{i=1}^{n} a^{i-1} e_i \colon a \in \mathbb{R} \setminus \{0\} \right\}.$$

Then any n vectors in W constitute a basis of $\mathbb{R}^n$ since their coefficients from a Vandermonde determinant. Since W is uncountable and $\mathcal{Q}$ is countable, there exists at least one $A \in \mathcal{Q}$ such that $|\{x \in W\colon Bx = Ax\}| \geqslant n$. Hence $B = A$ on a basis of $\mathbb{R}^n$, and so $B = A$ on $\mathbb{R}^n$.

Also solved by V. Anantharam, Carlos Biasi and Wilson Mauricio Tadini (Brazil), Jozef Bobok (Czechoslovakia), Jerzy Browkin (Poland), F. S. Cater, Chico Problem Group, John Cobb, G. Crofts, Patrick R. Gardner, Jaroslav Guričan (Czechoslovakia), William H. Gustafson, Dinh Thế Hưng and Benny Cheng, Elgin Johnston, Joel Levy, J. G. Mauldon, William Myers, Ivan Netuka (Czechoslovakia), Nicholas Passell, Eero Posti (Finland), Adam Riese, C. Ray Rosentrater, A. R. Sourour (Canada), R. A. Vitale, David M. Wells, J. B. Wilker (Canada), Pei Yuan Wu (Taiwan), and the proposers.

Separation of Edges of a Square by a Closed Set

6351 [1981, 540]. *Proposed by James W. Walker, MIT.*

Let F be a closed subset of the unit square $I^2 = [0, 1] \times [0, 1]$ such that F contains the closed horizontal edges $[0, 1] \times \{0\}$ and $[0, 1] \times \{1\}$, but does not intersect the open vertical edges $\{0\} \times (0, 1)$ and $\{1\} \times (0, 1)$. Show that either both horizontal edges belong to the same connected

component of F, or else both vertical edges belong to the same component of $I^2 \setminus F$.

Solution by John Cobb, University of Idaho. Suppose no component of $I^2 \setminus F$ contains both vertical edges. Then F separates these two edges and hence, by the unicoherence of I^2, some component G of F separates them also. (See M. H. A. Newman, *Elements of the Topology of Plane Sets of Points*, (Cambridge University Press), 2nd ed., Theorem 14.3.) If G fails to contain one of the horizontal edges, say the top, then the connected set

$$\{0\} \times (0,1) \cup [0,1] \times \{1\} \cup \{1\} \times (0,1)$$

lies in $I^2 \setminus G$, contradicting the separation of the vertical edges by G.

Also solved by the Chico Problem Group, Jay Hook, Terry L. Lay, Mark D. Meyerson, and the proposer.

Can S_n be Embedded in A_{n+1}?

6353 [1981, 541]. *Proposed by the editors.*

"Everyone" knows that the symmetric group S_n can be embedded in A_{n+2}. Can S_n be embedded in A_{n+1} (alternating group)?

I. *Solution by Lenny Jones, Shippensburg State College, Shippensburg, Pennsylvania.* Yes, for $n = 1$, and no, for $n > 1$. For $n = 1$ this is obvious. For $n \geqslant 2$, by way of contradiction, assume S_n can be embedded in A_{n+1}. Then $[A_{n+1}: S_n] = (n+1)/2$ which yields a contradiction unless n is odd and greater than 3. (The case $n = 3$ is disposed of using the "well-known" fact that A_4 has no subgroup of order 6.) The action of A_{n+1} on the cosets of S_n gives a homomorphism from A_{n+1} into $S_{(n+1)/2}$. This homomorphism has a nontrivial kernel since $(n+1)!/2 > ((n+1/2)!$ for odd $n \geqslant 5$. But this contradicts the fact that A_{n+1} is simple for $n \geqslant 5$.

II. *Solution by James Riddell, University of Victoria, Canada.* First consider cases $n \geqslant 13$. Such an n can be expressed $n = 2 + p + q + \cdots$ where $p, q, \ldots$ are distinct odd primes, since every integer $\geqslant 10$ can be partitioned into distinct odd primes (see Sierpinski, "Elementary Theory of Numbers," Warsaw, 1964, p. 144, Theorem 12). Then S_n contains a permutation of order $2 \cdot p \cdot q \cdots$, namely the product of disjoint cycles of the respective lengths $2, p, q, \ldots$. But A_{n+1} cannot contain a permutation of this order, since otherwise $n+1$ would have to be expressible $n + 1 = 2 + 2 + p + q + \cdots$.

For smaller n we need consider only n odd, since for n even, $n! \nmid (n+1)!/2$. The cases $n = 5 = 2 + 3$, $n = 7 = 2 + 5$, $n = 9 = 2 + 7$, and $n = 11 = 2 + 9$ can be argued similarly to the above, and the case $n = 3$ by noticing that S_3 has two elements of period 2 whose product has period 3 (and $\mathcal{C}_4$ does not have such a pair).

Also solved by D. M. Bloom, Duane M. Broline, Jerzy Browkin (Poland), D. W. Brown, Paul L. Chabot, Lorraine L. Foster, Robert Gilmer, James Henle & Stan Wagon, A. A. Jagers (The Netherlands), Michael Josephy (Costa Rica), John Kahila, Eugene Levine, Desmond MacHale (Ireland), Joseph J. Rotman, Claudia Spiro, Rony Teitler (England), Gary L. Walls, and Michael Woltermann. Partially solved by Peter R. Atwood and Kenneth Klinger.

MISCELLANEA

91. If I had been *trained* in research—that ridiculous contradiction in terms—should I have done more than produce additions to the existing store of little papers with blunted conclusions, of which there are already too many?

–"George Ponderevo" in H. G. Wells, *Tono-Bungay*, Book 2, Chapter 2.

component of F, or else both vertical edges belong to the same component of $I^2 \setminus F$.

Solution by John Cobb, University of Idaho. Suppose no component of $I^2 \setminus F$ contains both vertical edges. Then F separates these two edges and hence, by the unicoherence of I^2, some component G of F separates them also. (See M. H. A. Newman, *Elements of the Topology of Plane Sets of Points*, (Cambridge University Press), 2nd ed., Theorem 14.3.) If G fails to contain one of the horizontal edges, say the top, then the connected set

$$\{0\} \times (0,1) \cup [0,1] \times \{1\} \cup \{1\} \times (0,1)$$

lies in $I^2 \setminus G$, contradicting the separation of the vertical edges by G.

Also solved by the Chico Problem Group, Jay Hook, Terry L. Lay, Mark D. Meyerson, and the proposer.

Can S_n be Embedded in A_{n+1}?

6353 [1981, 541]. *Proposed by the editors.*

"Everyone" knows that the symmetric group S_n can be embedded in A_{n+2}. Can S_n be embedded in A_{n+1} (alternating group)?

I. *Solution by Lenny Jones, Shippensburg State College, Shippensburg, Pennsylvania.* Yes, for $n = 1$, and no, for $n > 1$. For $n = 1$ this is obvious. For $n \geqslant 2$, by way of contradiction, assume S_n can be embedded in A_{n+1}. Then $[A_{n+1}: S_n] = (n+1)/2$ which yields a contradiction unless n is odd and greater than 3. (The case $n = 3$ is disposed of using the "well-known" fact that A_4 has no subgroup of order 6.) The action of A_{n+1} on the cosets of S_n gives a homomorphism from A_{n+1} into $S_{(n+1)/2}$. This homomorphism has a nontrivial kernel since $(n+1)!/2 > ((n+1/2)!$ for odd $n \geqslant 5$. But this contradicts the fact that A_{n+1} is simple for $n \geqslant 5$.

II. *Solution by James Riddell, University of Victoria, Canada.* First consider cases $n \geqslant 13$. Such an n can be expressed $n = 2 + p + q + \cdots$ where $p, q, \ldots$ are distinct odd primes, since every integer $\geqslant 10$ can be partitioned into distinct odd primes (see Sierpinski, "Elementary Theory of Numbers," Warsaw, 1964, p. 144, Theorem 12). Then S_n contains a permutation of order $2 \cdot p \cdot q \cdots$, namely the product of disjoint cycles of the respective lengths $2, p, q, \ldots$. But A_{n+1} cannot contain a permutation of this order, since otherwise $n+1$ would have to be expressible $n + 1 = 2 + 2 + p + q + \cdots$.

For smaller n we need consider only n odd, since for n even, $n! \nmid (n+1)!/2$. The cases $n = 5 = 2 + 3$, $n = 7 = 2 + 5$, $n = 9 = 2 + 7$, and $n = 11 = 2 + 9$ can be argued similarly to the above, and the case $n = 3$ by noticing that S_3 has two elements of period 2 whose product has period 3 (and $\mathcal{Q}_4$ does not have such a pair).

Also solved by D. M. Bloom, Duane M. Broline, Jerzy Browkin (Poland), D. W. Brown, Paul L. Chabot, Lorraine L. Foster, Robert Gilmer, James Henle & Stan Wagon, A. A. Jagers (The Netherlands), Michael Josephy (Costa Rica), John Kahila, Eugene Levine, Desmond MacHale (Ireland), Joseph J. Rotman, Claudia Spiro, Rony Teitler (England), Gary L. Walls, and Michael Woltermann. Partially solved by Peter R. Atwood and Kenneth Klinger.

MISCELLANEA

91. If I had been *trained* in research—that ridiculous contradiction in terms—should I have done more than produce additions to the existing store of little papers with blunted conclusions, of which there are already too many?

–"George Ponderevo" in H. G. Wells, *Tono-Bungay*, Book 2, Chapter 2.

component of F, or else both vertical edges belong to the same component of $I^2 \setminus F$.

Solution by John Cobb, University of Idaho. Suppose no component of $I^2 \setminus F$ contains both vertical edges. Then F separates these two edges and hence, by the unicoherence of I^2, some component G of F separates them also. (See M. H. A. Newman, *Elements of the Topology of Plane Sets of Points*, (Cambridge University Press), 2nd ed., Theorem 14.3.) If G fails to contain one of the horizontal edges, say the top, then the connected set

$$\{0\} \times (0,1) \cup [0,1] \times \{1\} \cup \{1\} \times (0,1)$$

lies in $I^2 \setminus G$, contradicting the separation of the vertical edges by G.

Also solved by the Chico Problem Group, Jay Hook, Terry L. Lay, Mark D. Meyerson, and the proposer.

Can S_n be Embedded in A_{n+1}?

6353 [1981, 541]. *Proposed by the editors.*

"Everyone" knows that the symmetric group S_n can be embedded in A_{n+2}. Can S_n be embedded in A_{n+1} (alternating group)?

I. *Solution by Lenny Jones, Shippensburg State College, Shippensburg, Pennsylvania.* Yes, for $n = 1$, and no, for $n > 1$. For $n = 1$ this is obvious. For $n \geqslant 2$, by way of contradiction, assume S_n can be embedded in A_{n+1}. Then $[A_{n+1}: S_n] = (n+1)/2$ which yields a contradiction unless n is odd and greater than 3. (The case $n = 3$ is disposed of using the "well-known" fact that A_4 has no subgroup of order 6.) The action of A_{n+1} on the cosets of S_n gives a homomorphism from A_{n+1} into $S_{(n+1)/2}$. This homomorphism has a nontrivial kernel since $(n+1)!/2 > ((n + 1/2)!$ for odd $n \geqslant 5$. But this contradicts the fact that A_{n+1} is simple for $n \geqslant 5$.

II. *Solution by James Riddell, University of Victoria, Canada.* First consider cases $n \geqslant 13$. Such an n can be expressed $n = 2 + p + q + \cdots$ where $p, q, \ldots$ are distinct odd primes, since every integer $\geqslant 10$ can be partitioned into distinct odd primes (see Sierpinski, "Elementary Theory of Numbers," Warsaw, 1964, p. 144, Theorem 12). Then S_n contains a permutation of order $2 \cdot p \cdot q \cdots$, namely the product of disjoint cycles of the respective lengths $2, p, q, \ldots$. But A_{n+1} cannot contain a permutation of this order, since otherwise $n+1$ would have to be expressible $n + 1 = 2 + 2 + p + q + \cdots$.

For smaller n we need consider only n odd, since for n even, $n! \nmid (n+1)!/2$. The cases $n = 5 = 2 + 3$, $n = 7 = 2 + 5$, $n = 9 = 2 + 7$, and $n = 11 = 2 + 9$ can be argued similarly to the above, and the case $n = 3$ by noticing that S_3 has two elements of period 2 whose product has period 3 (and $\mathcal{A}_4$ does not have such a pair).

Also solved by D. M. Bloom, Duane M. Broline, Jerzy Browkin (Poland), D. W. Brown, Paul L. Chabot, Lorraine L. Foster, Robert Gilmer, James Henle & Stan Wagon, A. A. Jagers (The Netherlands), Michael Josephy (Costa Rica), John Kahila, Eugene Levine, Desmond MacHale (Ireland), Joseph J. Rotman, Claudia Spiro, Rony Teitler (England), Gary L. Walls, and Michael Woltermann. Partially solved by Peter R. Atwood and Kenneth Klinger.

MISCELLANEA

91. If I had been *trained* in research—that ridiculous contradiction in terms—should I have done more than produce additions to the existing store of little papers with blunted conclusions, of which there are already too many?

–"George Ponderevo" in H. G. Wells, *Tono–Bungay*, Book 2, Chapter 2.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Topologie. Fonctions réelles d'une variable réelle. Second Edition. By A. Doneddu. (Nouveau cours de mathématiques, vol. 4.) Vuibert, Paris, France, 1979. 315 pp.

R. P. BOAS
Department of Mathematics, Northwestern University, Evanston, IL 60201

We all know that Mathematics has changed greatly during the last half-million class periods—fifty years, that is. It has grown at an accelerated rate, and many important problems have been solved in our time after resisting many efforts. Has Mathematics Education kept up?

Well, to some extent. Half a century ago I was a student in a course, more or less at the level of Doneddu's book. The instructor had been specifically warned by the Department not to mention uniform convergence (it was too difficult a concept!). Doneddu does discuss uniform convergence, in fact, in exactly the terms in which I was to encounter it the next year (although I think, now, that there are better ways). In half a century we have gained perhaps a year. Shouldn't we expect more than that from a field that is growing explosively?

Doneddu says that his book was written as one of a series of texts prescribed by a decree of the French government in 1972. The other volumes are: (1) *Structures fondamentales*, (2) *Polynômes et algèbre linéaire*, (3) *Espaces euclidiens et hermitiens*, *Géometries*, (5) *Fonctions Vectorielles*, *Séries*, *Equations différentielles*, (6) *Géométrie différentielle*, *Intégrales multiples*. Give or take a few topics, this could be an outline of an advanced calculus course of the 1930's, except for the first part of volume 4. I presume that these volumes are not intended as a first course for beginners, although technically they could be. (In fact I know a professional mathematician who taught himself calculus from a three-volume *Vorlesungen über Differential-und Integralrechnung*.) Volume 4 could be used as a handy reference in which you could look up concise and reliable statements and proofs of calculus theorems. But be warned: Doneddu's exposition suffers from his apparent belief that more terminology means more rigor. Presumably this will not trouble the intended audience, who must have been trained at Bourbaki's knee, but it does create trouble for outsiders. For example, Doneddu provides no index entries for some of Bourbaki's favorite words and no index of symbols. I had to turn a good many pages, one by one, before I found out for sure what $\mathfrak{F}(E, R)$ stood for.

If the students who use this book ever learn that calculus is used for something, whether in or out of mathematics, they will not learn it from the book itself. Considering that the directions taken by calculus have been largely determined by its applications to geometry, physical sciences, and more recently to social sciences, this omission of applications seems hard to justify. Calculus is too close to the real world to be presented as an abstract game.

It is clearly too much to hope that textbooks on calculus, at whatever level, will start including very much new mathematics very soon, but they could at least show some awareness of what has been going on. Doneddu, for example, mentions computers and calculators in passing, but gives no indication of the influence that these tools have had on many parts of mathematics, and on numerical analysis in particular. Indeed, he has a chapter on numerical analysis that could have been written 50 years ago, except that it is much less sophisticated than what would have been acceptable then.

Anyone ought to be able to think of a dozen significant topics at the advanced calculus level that could reasonably count as advanced calculus, but are represented in few, if any, current textbooks on advanced calculus (or introductory real analysis, or whatever it happens to be called). (My own list would include generalized functions (distributions), constrained extremal

problems with inequality constraints, Abel's theorem for power series, inequalities, and methods of approximating infinite series and integrals.) Are such topics omitted because they are not traditional? They do not appear in books written for engineers and scientists, either; is this because the corresponding faculties themselves use only more traditional mathematics? I am rather afraid so.

You are probably going to ask how more topics can be put into a course in advanced calculus that is already too crowded. My guess is that it is too crowded because people are trying to tell everything that they know about each topic. It is harder, but more rewarding, to try to pare each topic down to essentials, and then cover more topics.

Pathways to Solutions, Fixed Points, and Equilibria. By C. B. Garcia and W. I. Zangwill. Prentice-Hall, Englewood Cliffs, New Jersey, 1981. xvi + 479 pp., $32.00.

WERNER C. RHEINBOLDT
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15261

For over a century the continuation principle has proved to be an important tool in mathematics. To illustrate the idea, suppose that for a point $y(0)$ on some manifold M_y exactly one point $x(0)$ with a specific property is known to exist on another manifold M_x. Then the same assertion may be shown to hold for another point $y(1)$ of M_y by using continuity arguments to prove it for all points $y(s)$, $0 \leqslant s \leqslant 1$, of a path on M_y connecting $y(0)$ with $y(1)$. Early uses of this approach date at least to H. A. Schwarz and his work on conformal mappings around 1869. Then J. Hadamard and M. Levy applied it in connection with the inversion of nonlinear mappings, and it is also a basic tool in J. Leray's and J. Schauder's celebrated work on differential equations. These are only a few of many possible examples.

Until a few decades ago the continuation principle was strictly considered a theoretical tool; for example, there would have been no interest in computing explicitly the paths in our sketchy illustration. The first application of the approach to a numerical problem appears to go back to a little noticed paper by E. Lahaye in 1932, but it was not until the early fifties and the advent of automatic computing that continuation techniques began to be used widely for numerical work.

One of the main problems to which these techniques are applied concerns the solution of a nonlinear equation $Fx = 0$ in R^n. In order to compute a specific solution x^*, a typical approach is to imbed the given equation into a family of equations $H(x, t) = 0$, $0 \leqslant t \leqslant 1$, for which there exists a computable solution path $x = x(t)$, $0 \leqslant t \leqslant 1$, beginning at a known point $x(0)$ and ending at $x(1) = x^*$. A related but conceptually distinct problem arises in many areas of engineering, notably structural mechanics, where the equation under consideration always includes a number of physically important, intrinsic parameters. In these problems it is of little interest to determine only a few specific solutions. Instead, the requirement is to follow paths on the solution-manifold in the space of all state and parameter variables and thereby to detect specific features of the manifold which signify a change of behavior of the system under study.

For the first of these problems, namely the computation of a specific solution of some equation, two distinct continuation techniques are available. The first one involves the case when the solution paths are differentiable which in turn allows their representation as a solution of a differential equation. The second approach was initiated by H. Scarf in 1967 in a very different setting. In fact, his technique is based on homological rather than homotopic concepts and—as observed by H. Kuhn—made use of a numerical form of Sperner's lemma. Since then this

problems with inequality constraints, Abel's theorem for power series, inequalities, and methods of approximating infinite series and integrals.) Are such topics omitted because they are not traditional? They do not appear in books written for engineers and scientists, either; is this because the corresponding faculties themselves use only more traditional mathematics? I am rather afraid so.

You are probably going to ask how more topics can be put into a course in advanced calculus that is already too crowded. My guess is that it is too crowded because people are trying to tell everything that they know about each topic. It is harder, but more rewarding, to try to pare each topic down to essentials, and then cover more topics.

Pathways to Solutions, Fixed Points, and Equilibria. By C. B. Garcia and W. I. Zangwill. Prentice-Hall, Englewood Cliffs, New Jersey, 1981. xvi + 479 pp., $32.00.

WERNER C. RHEINBOLDT
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15261

For over a century the continuation principle has proved to be an important tool in mathematics. To illustrate the idea, suppose that for a point $y(0)$ on some manifold M_y exactly one point $x(0)$ with a specific property is known to exist on another manifold M_x. Then the same assertion may be shown to hold for another point $y(1)$ of M_y by using continuity arguments to prove it for all points $y(s)$, $0 \leqslant s \leqslant 1$, of a path on M_y connecting $y(0)$ with $y(1)$. Early uses of this approach date at least to H. A. Schwarz and his work on conformal mappings around 1869. Then J. Hadamard and M. Levy applied it in connection with the inversion of nonlinear mappings, and it is also a basic tool in J. Leray's and J. Schauder's celebrated work on differential equations. These are only a few of many possible examples.

Until a few decades ago the continuation principle was strictly considered a theoretical tool; for example, there would have been no interest in computing explicitly the paths in our sketchy illustration. The first application of the approach to a numerical problem appears to go back to a little noticed paper by E. Lahaye in 1932, but it was not until the early fifties and the advent of automatic computing that continuation techniques began to be used widely for numerical work.

One of the main problems to which these techniques are applied concerns the solution of a nonlinear equation $Fx = 0$ in R^n. In order to compute a specific solution x^*, a typical approach is to imbed the given equation into a family of equations $H(x, t) = 0$, $0 \leqslant t \leqslant 1$, for which there exists a computable solution path $x = x(t)$, $0 \leqslant t \leqslant 1$, beginning at a known point $x(0)$ and ending at $x(1) = x^*$. A related but conceptually distinct problem arises in many areas of engineering, notably structural mechanics, where the equation under consideration always includes a number of physically important, intrinsic parameters. In these problems it is of little interest to determine only a few specific solutions. Instead, the requirement is to follow paths on the solution-manifold in the space of all state and parameter variables and thereby to detect specific features of the manifold which signify a change of behavior of the system under study.

For the first of these problems, namely the computation of a specific solution of some equation, two distinct continuation techniques are available. The first one involves the case when the solution paths are differentiable which in turn allows their representation as a solution of a differential equation. The second approach was initiated by H. Scarf in 1967 in a very different setting. In fact, his technique is based on homological rather than homotopic concepts and—as observed by H. Kuhn—made use of a numerical form of Sperner's lemma. Since then this

approach has been reformulated and is now considered principally in the form of algorithms involving piecewise-linear solution paths.

Up to now most of this material on the numerical analysis of both of these forms of continuation algorithms could be found mainly in the journal literature or in conference proceedings. This book represents a first introduction to the area on a level accessible to graduate students and to all those interested in applying these methods in their own fields.

The authors' own area of interest centers around applications in mathematical programming and economics. It is therefore not surprising that the corresponding chapters are probably the strongest ones in the book. There are very nice discussions of applications to nonlinear programming, economic equilibria, and game theory. Correspondingly, the presentations of the piecewise linear algorithms and of some of the underlying concepts, such as linear complementarity, are especially well done.

In contrast, the case of differentiable paths is dealt with in a somewhat less rounded form, and, in fact, the discussion of the corresponding algorithms seems rather sketchy. There is no consideration of the indicated problem of using continuation algorithms for analyzing solution manifolds, and very little mention is made of the extensive engineering literature in this area. Only a few elementary examples of structural equilibrium problems are included. Similarly, the brief discussion of the ideas behind catastrophe theory only skirts the central points of the theory, and the considerations of the loss of stability of the equilibria under study are not related to any numerical algorithms.

However, these last observations should not detract from the overall value of the book. In fact, any book which attempts to cover a fairly large area with relatively elementary mathematical techniques cannot satisfy every reader. This text is indeed a very nice introduction to the field which should serve its intended purpose well, namely, to make this circle of concepts and techniques accessible to a wider audience, and to provide a possible text for graduate courses and seminars on these topics. For this it deserves a warm recommendation.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

7 June 1982

Editor:

I am, and have been since 1962, *the* mathematics department at this small (student body about 400) "liberal arts" college. My normal teaching load is anywhere from 14 to 17 hours each semester—4 to 5 different math classes. As a result, and I suppose by choice, my mathematical involvement is other than research. I happen to like (math) teaching, and, even more, I am most happy doing or reading some sort of mathematics.

About 80% of my math majors go on to graduate school. My course work for them involves everything from calculus to topology, complex variables, and operations research.

All this is prelude to justify what *I* want in the MONTHLY; namely "folklore" ideas in various fields. Examples: proving the vector product distributes over addition by pseudo-division, an accessible (to junior and senior math majors) topological discussion of orientation relative to Gauss' theorem (manifold theory), use of the Sylow theorems and other techniques in showing the

approach has been reformulated and is now considered principally in the form of algorithms involving piecewise-linear solution paths.

Up to now most of this material on the numerical analysis of both of these forms of continuation algorithms could be found mainly in the journal literature or in conference proceedings. This book represents a first introduction to the area on a level accessible to graduate students and to all those interested in applying these methods in their own fields.

The authors' own area of interest centers around applications in mathematical programming and economics. It is therefore not surprising that the corresponding chapters are probably the strongest ones in the book. There are very nice discussions of applications to nonlinear programming, economic equilibria, and game theory. Correspondingly, the presentations of the piecewise linear algorithms and of some of the underlying concepts, such as linear complementarity, are especially well done.

In contrast, the case of differentiable paths is dealt with in a somewhat less rounded form, and, in fact, the discussion of the corresponding algorithms seems rather sketchy. There is no consideration of the indicated problem of using continuation algorithms for analyzing solution manifolds, and very little mention is made of the extensive engineering literature in this area. Only a few elementary examples of structural equilibrium problems are included. Similarly, the brief discussion of the ideas behind catastrophe theory only skirts the central points of the theory, and the considerations of the loss of stability of the equilibria under study are not related to any numerical algorithms.

However, these last observations should not detract from the overall value of the book. In fact, any book which attempts to cover a fairly large area with relatively elementary mathematical techniques cannot satisfy every reader. This text is indeed a very nice introduction to the field which should serve its intended purpose well, namely, to make this circle of concepts and techniques accessible to a wider audience, and to provide a possible text for graduate courses and seminars on these topics. For this it deserves a warm recommendation.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

7 June 1982

Editor:

I am, and have been since 1962, *the* mathematics department at this small (student body about 400) "liberal arts" college. My normal teaching load is anywhere from 14 to 17 hours each semester—4 to 5 different math classes. As a result, and I suppose by choice, my mathematical involvement is other than research. I happen to like (math) teaching, and, even more, I am most happy doing or reading some sort of mathematics.

About 80% of my math majors go on to graduate school. My course work for them involves everything from calculus to topology, complex variables, and operations research.

All this is prelude to justify what *I* want in the MONTHLY; namely "folklore" ideas in various fields. Examples: proving the vector product distributes over addition by pseudo-division, an accessible (to junior and senior math majors) topological discussion of orientation relative to Gauss' theorem (manifold theory), use of the Sylow theorems and other techniques in showing the

actual construction of non-isomorphic subgroups of groups of order 10 to 17, and developments in operations research explained in less than 500 pages.

I do like the history of development of mathematical concepts. Think the students really profit from learning about the difficult times the "early day" people had in smoothing out the ideas so glibly presented in our short class periods. I like to use the concept of the evolution of normal subgroups in this context.

Book reviews are important to me. Publishers do not have much to do with small colleges. How my colleagues react to current texts is important to me. Examples: I switched from Thomas & Finney to Philip Gilletts calculus, I use Boas' *A Primer of Real Functions* in my one semester introduction to real variables course—how do others react to these texts? I refuse to teach a course in computer programming, yet I have my students learn (by some instruction in first year calculus) how to program and run the Hewlett-Packard 41. I feel as if any more is not too productive—am I way off base?

William H. Jamison
Physics and Mathematics Department
Rocky Mountain College
Billings, Montana 59102

Comments, anyone? Ed.

Letter *from* the editor.

One complaint about the MONTHLY keeps recurring to almost every editor almost every year: the articles are too hard, too technical, too specialized. Perhaps only a few think that, but, if so, they are vocal out of proportion to their numbers; if most readers (possible readers) think that, something ought to be done about it.

I reject research papers out of hand; the MONTHLY is not a research outlet. If the referees (and my own instincts) tell me that a would-be expository paper is just not expository enough, many people (referees, authors, associate editors, editorial assistants) go to work and try to prepare a good revision.

What else can be done? An editor doesn't write the articles. I sometimes solicit them, but mostly I just have to wait till they come in. I would like to hear from the readers: do you agree with the complaint, and do you have a suggestion for a cure? Would you be willing to write the proper kind of article—because you have something to say, because you can say something better than the last person who tried it, or just to show that it can be done?

Paul Halmos

Developmental Delivery Systems That Deliver

A diagnostic/prescriptive approach. . .

PCSP: Personalized Computational Skills Program
A Brush-Up Book®
Bryce R. Shaw, Senior Author
Gwen Shelton, Bee County College, Texas
Sandra Pryor Clarkson, Hunter College, New York

VOLUME 1 COMPUTATIONAL SKILLS • 1980
Module A: Whole Numbers
Module B: Fractions
Module C: Decimals, Proportions, and Percents

VOLUME 2 APPLICATIONS • 1981
Module D: Applications Using Computational Skills
Module E: Special Skills with Applications
Module F: Beginning Algebra with Applications

PCSP COMBINED EDITION • 1982

PCSP uses skill models, worked examples, and thousands of practice problems and applications to improve computational competency. Built-in diagnostic and mastery tests pinpoint students' problem areas and measure their learning progress.

The six modules are available separately or in two volumes, which may be used independently, concurrently, or in sequence. The Combined Edition offers a quick brush-up on basic skills together with their applications in a single volume.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 90, Number 2 February 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

AWARD FOR DISTINGUISHED SERVICE TO PROFESSOR EDWIN F. BECKENBACH

IVAN NIVEN
Department of Mathematics, University of Oregon, Eugene OR 97403

This year's recipient of the Award for Distinguished Service to Mathematics was a widely-known member of the mathematical community, active for many decades in research, in teaching, in expository writing, and in dedicated work on behalf of professional organizations. Today we honor Edwin F. Beckenbach for the long and outstanding service he gave in such a variety of ways. His influence has been widespread, significant, and lasting.

Born in Dallas in 1906, Ed Beckenbach attended school there, graduating as valedictorian of his high school class. In the fall of 1924 Ed entered Rice University in Houston, where he continued to his Ph.D. in 1931. His dissertation was written under the direction of Lester R. Ford, Sr., a distinguished mathematician who served as editor of the MONTHLY some forty years ago. Tibor Radó also strongly influenced his work.

The depression year 1931 was not an auspicious time to move into the job market, but Ed survived those tough times of the 1930's very well. After two years as a National Research Fellow at Princeton, Ohio State, and Chicago, he accepted a faculty position at Rice University. He moved on later to the University of Michigan, then to the University of Texas, and finally to UCLA in 1945, where he served for three decades. Ed Beckenbach is one of the scholars who brought UCLA into the national prominence it now enjoys. Among other contributions, Ed played a central role in the establishment of the Institute for Numerical Analysis there.

Ed wrote extensively on inequalities, on geometric complex variable theory, and on its extension, through subharmonic functions, to the theory of surfaces. For example, he has written two books on inequalities with Richard Bellman, one in the New Mathematical Library Series of the Association, and a more advanced book which has become a standard research reference work. He edited *Modern Mathematics for the Engineer*, vols. 1 and 2, *Applied Combinatorial Mathematics*, and *Concepts of Communication*, with various translations into eight foreign languages, and is coeditor of a mathematics dictionary. In addition to these advanced works, Ed is author, or coauthor of several books at the elementary, intermediate, and collegiate levels. He served on writing teams for the School Mathematics Study Group, the National Council of Teachers of Mathematics, and the African Education Program. In addition to all this writing and editing, Ed contributed his time and talent to professional organizations, including of course the Association.

He was a member of the Council of the American Mathematical Society, and chairman of a special blue ribbon Committee on Mathematical Reviews at a time when MR was in considerable distress. He served on the membership committee of the Society for Industrial and Applied Mathematics, and was chairman of a special committee to reorganize the election procedures of the Conference Board of the Mathematical Sciences. Ed was a trustee of the Association of Members of the Institute for Advanced Study, and chairperson of its Western Committee. He was one of the organizers of a week-long conference on mathematical inequalities, held biennially at Oberwolfach in Germany. He edited the Proceedings of these international conferences.

Lengthy listings tend to become dreary, so this report is confined to a sample of Ed Beckenbach's many and varied contributions to the world of mathematics. No report would be complete, however, without drawing attention to Ed's role in the founding and editing of the *Pacific Journal of Mathematics*. Indeed, the January 1982 issue of that journal is dedicated to the two pioneers who were most instrumental in getting the enterprise off the ground, Ed Beckenbach and Frantisek Wolf of Berkeley. Few of us play central roles in the founding of a journal. Fewer still have the vision, the dedication and the energy to launch not one but two journals, and Ed is one of this distinguished group. Ed's urging over many years that the Association should have a

EDWIN F. BECKENBACH, 1906–1982

newsletter culminated in the creation of our very successful, very well-received newsletter, *FOCUS*. Its editor, Marcia P. Sward, calls Ed the "Father of *FOCUS*".

Let me conclude by emphasizing other service to the Association. Ed was elected chairman of the Texas Section, but he left for Michigan before he took office. He was a visiting lecturer for the MAA, and served a five-year term as editor of the Mathematical Notes Section of the MONTHLY. As chairman of the Committee on Publications since 1971, he guided the unprecedented growth of our journals and our series of books and monographs with skill, determination, and enthusiasm.

In all his activities, Ed Beckenbach enlisted the cooperation of his colleagues by his skill at negotiation, his unfailing courtesy and consideration toward others, and his common sense and good humor. But Ed's cooperative and accommodating spirit at the committee table completely disappeared in another of his roles. On the tennis court Ed was a hard contender, a tough adversary who showed 'em no mercy. Captain of the tennis team at Rice University back in the twenties, and later the coach of the team, he spanned six decades with his favorite sport. Even recently Ed and his wife Alice competed in national tournaments of "superseniors"; in more peaceful moments they indulged their lifelong hobby, tending their hillside acre of plants ranging from apricots to orchids. Incidentally, Alice surely holds a record for faithful attendance at meetings of mathematicians, having started at the age of eleven as daughter of a one-time president of the Association.

Ed Beckenbach clearly served the mathematical community very well. We are all indebted to him for his preeminent leadership. Ed Beckenbach is indeed a most worthy recipient of the Award for Distinguished Service to Mathematics.

* * *

Edwin F. Beckenbach died on September 5, 1982.

SINGULAR VALUE ANALYSIS OF CRYPTOGRAMS

CLEVE MOLER AND DONALD MORRISON*

Department of Computer Science, University of New Mexico, Albuquerque, NM 87131

1. Singular Value Analysis and Cryptanalysis. The singular value decomposition is a matrix factorization which can produce approximations to large arrays. Cryptanalysis is the task of breaking coded messages. In this paper, we present an unusual merger of the two in which the singular value decomposition may aid the cryptanalyst in discovering vowels and consonants in messages coded in certain variations of simple substitution ciphers.

Texts in many languages, including English, have the property that vowels are frequently followed by consonants, and consonants are frequently followed by vowels. We say a text is a

Cleve B. Moler received his Ph.D. at Stanford under George Forsythe. He has taught both mathematics and computer science at Stanford, the University of Michigan, and the University of New Mexico. He recently succeeded Morrison as Chairman of the Department of Computer Science at New Mexico. His academic and research interests include: numerical analysis, mathematical software, and scientific computing.

Don Morrison earned his Ph.D. in mathematics at the University of Wisconsin in 1950, under C. C. MacDuffee. He taught mathematics at Tulane, 1950–55, and computer science at the University of New Mexico, 1967 to the present. He was a staff member, division supervisor and department manager at Sandia Corporation, 1955–71. He has one wife, three children and numerous grandchildren. His present research interests include cryptology, data structures, and design of algorithms.

*This work was partially supported by a grant from the National Science Foundation.

newsletter culminated in the creation of our very successful, very well-received newsletter, *FOCUS*. Its editor, Marcia P. Sward, calls Ed the "Father of *FOCUS*".

Let me conclude by emphasizing other service to the Association. Ed was elected chairman of the Texas Section, but he left for Michigan before he took office. He was a visiting lecturer for the MAA, and served a five-year term as editor of the Mathematical Notes Section of the MONTHLY. As chairman of the Committee on Publications since 1971, he guided the unprecedented growth of our journals and our series of books and monographs with skill, determination, and enthusiasm.

In all his activities, Ed Beckenbach enlisted the cooperation of his colleagues by his skill at negotiation, his unfailing courtesy and consideration toward others, and his common sense and good humor. But Ed's cooperative and accommodating spirit at the committee table completely disappeared in another of his roles. On the tennis court Ed was a hard contender, a tough adversary who showed 'em no mercy. Captain of the tennis team at Rice University back in the twenties, and later the coach of the team, he spanned six decades with his favorite sport. Even recently Ed and his wife Alice competed in national tournaments of "superseniors"; in more peaceful moments they indulged their lifelong hobby, tending their hillside acre of plants ranging from apricots to orchids. Incidentally, Alice surely holds a record for faithful attendance at meetings of mathematicians, having started at the age of eleven as daughter of a one-time president of the Association.

Ed Beckenbach clearly served the mathematical community very well. We are all indebted to him for his preeminent leadership. Ed Beckenbach is indeed a most worthy recipient of the Award for Distinguished Service to Mathematics.

* * *

Edwin F. Beckenbach died on September 5, 1982.

SINGULAR VALUE ANALYSIS OF CRYPTOGRAMS

CLEVE MOLER AND DONALD MORRISON*

Department of Computer Science, University of New Mexico, Albuquerque, NM 87131

1. Singular Value Analysis and Cryptanalysis. The singular value decomposition is a matrix factorization which can produce approximations to large arrays. Cryptanalysis is the task of breaking coded messages. In this paper, we present an unusual merger of the two in which the singular value decomposition may aid the cryptanalyst in discovering vowels and consonants in messages coded in certain variations of simple substitution ciphers.

Texts in many languages, including English, have the property that vowels are frequently followed by consonants, and consonants are frequently followed by vowels. We say a text is a

---

Cleve B. Moler received his Ph.D. at Stanford under George Forsythe. He has taught both mathematics and computer science at Stanford, the University of Michigan, and the University of New Mexico. He recently succeeded Morrison as Chairman of the Department of Computer Science at New Mexico. His academic and research interests include: numerical analysis, mathematical software, and scientific computing.

Don Morrison earned his Ph.D. in mathematics at the University of Wisconsin in 1950, under C. C. MacDuffee. He taught mathematics at Tulane, 1950–55, and computer science at the University of New Mexico, 1967 to the present. He was a staff member, division supervisor and department manager at Sandia Corporation, 1955–71. He has one wife, three children and numerous grandchildren. His present research interests include cryptology, data structures, and design of algorithms.

*This work was partially supported by a grant from the National Science Foundation.

"vowel-follows-consonant" text, or more briefly, a "vfc" text if the proportion of vowels following vowels is less than the proportion of vowels following consonants. That is, if

$$\frac{\text{number of vowel-vowel pairs}}{\text{number of vowels}} < \frac{\text{number of consonant-vowel pairs}}{\text{number of consonants}}. \tag{1}$$

Some languages produce better vfc texts than others and some deviations by individual letters can be expected. In English text, the letter h is often preceded by other consonants to form a single sound, as in ch, gh, ph, sh, and, especially th. The letters l, n, m, and r are often followed by consonants. Nevertheless, English is a predominantly vfc language. In Hawaiian, every consonant is followed by a vowel, so there are no consonant-consonant pairs. In the romaji transliteration of Japanese, the only consonant-consonant pairs are ch, sh, ts, and n, followed by a consonant and several double consonants. Russian has many sounds which, in transliteration, are of the consonant-consonant or vowel-vowel form, such as sh, ch, ts, ya, and ye, but in the Cyrillic alphabet, these are represented as single letters. Thus, Russian is also a predominantly vfc language.

2. The Cryptanalyst's Problem. A simple substitution cryptogram is a coded message in which the individual letters have been replaced by a permutation of themselves. When faced with such a message, a cryptanalyst might first count the occurrences of individual letters in the message and compare the frequencies with the known frequencies in typical uncoded text. However, a fairly long message is required before this technique has much chance of success.

Another aid in breaking the code is a partitioning of the alphabet of the cryptogram into two subsets, representing the vowels and the consonants. In order that such a partition be plausible, it ought to satisfy the vfc rule (1). This is the main task which we wish to consider here—and is what we call the cryptanalyst's problem. A trial and error solution, which tries all possible partitions until one satisfying the vfc rule is found, is clearly prohibitive.

Let n be the number of letters in the alphabet. For any text, the digram frequency matrix is the n-by-n array A with a_{ij} = the number of occurrences of the ith letter followed by the jth letter. Blanks and punctuation, if present, are ignored and the first letter of the text is assumed to follow the last letter. In general, the matrix is not symmetric, but for each i

$$\sum_j a_{ij} = \sum_j a_{ji} = f_i$$

where f_i = the number of occurrences of the ith letter.

For any proposed partitioning of the alphabet into vowels and consonants, two column vectors, v and c, can be defined by

$$v_i = 1 \text{ if the } i\text{th letter is a vowel, 0 otherwise,}$$
$$c_i = 1 \text{ if the } i\text{th letter is a consonant, 0 otherwise.}$$

Note that $v + c$ is a vector with all 1's and that the inner product v^Tc is zero. Also note that the value of the quadratic form v^TAv is the number of vowel-vowel pairs in the text.

Using A, v and c, the vfc rule (1) can be stated

$$\frac{v^TAv}{v^TA(v+c)} < \frac{c^TAv}{c^TA(v+c)}.$$

Cross-multiplying and cancelling the common term, we obtain

$$(v^TAv)(c^TAc) - (v^TAc)(c^TAv) < 0. \tag{2}$$

The cryptanalyst's problem, then, is: Given A, find a partitioning v and c so that (2) holds.

3. The Singular Value Decomposition. The singular value decomposition, or SVD, is a matrix factorization which numerical analysts use in a wide variety of ways. Although its primary uses are

in the analysis of systems of simultaneous linear equations and in the computation of pseudoinverses, we will use it here to obtain "simple" approximations to the digram frequency matrix. For this purpose, we express the SVD as a sum of rank one matrices of the form

$$A = \sigma_1 x_1 y_1^T + \sigma_2 x_2 y_2^T + \cdots + \sigma_n x_n y_n^T \tag{3}$$

where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

$$x_i^T x_j = \delta_{ij} \text{ (the Kronecker delta)}$$

$$y_i^T y_j = \delta_{ij}.$$

The coefficients σ_j are known as the *singular values* and x_j, and y_j are the *left* and *right* singular vectors, respectively. They can also be characterized in terms of the solutions to the symmetric eigenvalue problem:

$$\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

It is not hard to see that the *rank* of a matrix is the number of nonzero singular values. (In fact, this is a particularly useful way to define rank.)

There is a fast, reliable algorithm for computing the SVD [3], [6]. For a 26-by-26 matrix, the computation of the singular values and corresponding left and right vectors takes only a few seconds on a medium speed modern computer.

The normalization of the vectors x_j and y_j insures that all of the rank one matrices $x_j y_j^T$ have the same norm (that is, the sum of the squares of the elements). Consequently, the numerical importance of each term in the sum (3) can be measured by the size of the coefficient σ_j. If the σ_j decrease fairly rapidly as j increases, then an accurate approximation to A can be obtained by truncating the series after only a few terms. Truncating the series after k nonzero terms provides a rank k approximation to A. Such approximations are related to those obtained by factor analysis and have a wide variety of applications. One unusual application is to digital image processing [1].

4. Rank One Approximation. A rough approximation to a digram frequency matrix can be obtained by taking only the first term in its SVD:

$$A \approx A_1 = \sigma_1 x_1 y_1^T.$$

Let e be the vector of all 1's and let f be the vector with f_i = the number of occurrences of the ith letter in the text. Then $Ae = A^Te = f$ and so

$$\sigma_1 \left(y_1^T e \right) x_1 \approx \sigma_1 \left(x_1^T e \right) y_1 \approx f.$$

Consequently, if we were to assume that the digram frequency matrix were only rank one, we would conclude that it was symmetric and that each row and column was proportional to the frequency vector f.

Of course, in practice, A is not rank one. Nevertheless, the first left and right singular vectors tend to be approximately equal and reflect the frequencies of the letters in the text.

5. Rank Two Approximation. The rank two approximation obtained from the first two terms of the SVD is the simplest approximation which takes into account the correlation between pairs of letters in the text. The second left and right singular vectors contain the key to the solution of the cryptanalyst's problem.

Assume that A is approximated by a matrix of rank two, so that

$$A \approx A_2 = \sigma_1 x_1 y_1^T + \sigma_2 x_2 y_2^T. \tag{4}$$

We propose to use the signs of the components of x_2 and y_2 to partition the alphabet as follows:

$$v_i = \begin{cases} 1 & \text{if } x_{i2} > 0 \text{ and } y_{i2} < 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$(5) \qquad c_i = \begin{cases} 1 & \text{if } x_{i2} < 0 \text{ and } y_{i2} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$n_i = \begin{cases} 1 & \text{if sign}\,(x_{i2}) = \text{sign}\,(y_{i2}) \\ 0 & \text{otherwise.} \end{cases}$$

The third category—the "neuter" letters—are the ones that cannot be classified as either vowels or consonants. It turns out in practice that few letters fall into this category. In English text, for example, the letter h is usually neuter. It might be possible to consider a finer partition involving "left vowel, right consonant" and so on, but we have not pursued this idea.

Using these definitions, we obtain a solution to the cryptanalyst's problem as follows.

THEOREM. *Let $A = A_2$ be a nonnegative rank 2 matrix with the* SVD *expansion in* (4). *Let v and c be defined by* (5). *Then the* vfc *rule,*

$$(6) \qquad D = (v^TAv)(c^TAc) - (v^TAc)(c^TAv) < 0$$

is satisfied.

Proof. Since A is a nonnegative matrix, it follows from the Perron-Frobenius theorem [4] that x_1 and y_1 have nonnegative components. Placing (4) in (2) and expanding produces eight terms. The two terms involving only subscript 1 cancel, so do the two terms involving only subscript 2:

$$D = \sigma_1\sigma_2\big(v^Tz_1y_1^Tv\;c^Tz_2y_2^Tc + v^Tz_2y_2^Tv\;c_1^Tz\;y_1^Tc \\ - v^Tz_1y_1^Tc\;c^Tz_2y_2^Tv - v^Tz_2y_2^Tc\;c^Tz_1y_1^Tv\big).$$

Of all the different inner products appearing in this expression, only two—namely, y_2^Tv and c^Tz_2—are negative. Consequently, all four terms in the parentheses are negative and D is negative.

6. Effects of Encipherment. When a message M is encoded by a simple substitution cipher c, (where c is a permutation of the integers 1 to n) each occurrence of the ith letter u_i is replaced by the $c(i)$th letter, $u_{c(i)}$. The resulting cryptogram is called M_c. If A is the digram frequency matrix of M, and C is the permutation matrix $C = (\delta_{c(i),\,j})$ which has in its ith row, the $c(i)$th row of the identity matrix, then it is not difficult to show that the digram frequency matrix of M_c is CAC^T. This matrix has in row $c(i)$, column $c(j)$, the frequency of the digram u_iu_j in M, which it represents. It follows that the singular values of CAC^T are the same as those of A, and the jth left and right singular vectors are, respectively, Cx_j and Cy_j. These have the same coefficients as x_j and y_j, but they are permuted by C; the coefficients appearing in row i in x_j and y_j appear in row $c(i)$ in Cx_j and Cy_j. If the scheme described above, applied to A, classifies the letter u_i as a vowel in M, then applied to CAC^T, it will classify $u_{c(i)}$, the encoding of u_i, as a vowel in M_c.

Simple substitution ciphers are, of course, very simple ciphers, and no self-respecting cryptanalyst regards them as a challenge. A more sophisticated cipher is the k-alphabetic cipher (k is a positive integer). In this cipher, k permutations, $c_1, c_2, \ldots, c_k$ are used in cyclic fashion to encode the letters in M to produce the cryptogram M_c. A letter is said to be in position p, in M $(1 \le p < k)$, if it is the mth letter of M and m is congruent p modulo k. If the letter u_i occurs in position p, it is encoded as $u_{c_p(i)}$. Thus, the encoding of a letter depends, not only on what letter it is, but also on its position in M.

The digram frequency matrix of a cryptogram encoded by a k-alphabet cipher is of little use in decoding it, since the various occurrences of a digram in M_c do not represent the same digram in M, if they do not occur in the same position. Cryptanalysts have some clever ways to deduce the probable value of the cycle length k. See, for example, Gaines [2] or Sinkov [5]. Armed with this information, a cryptanalyst can calculate k digram frequency matrices, $A_p = (a_{ijp})$ where a_{ijp} is

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0	1	2	5	0	1	4	0	2	0	1	9	0	15	0	1	0	10	5	36	1	8	0	0	1	0
B	1	0	0	0	5	0	0	0	1	0	0	1	0	0	1	0	0	2	0	0	2	0	0	0	1	0
C	12	0	0	0	4	0	0	2	1	0	0	0	0	0	7	0	0	4	0	1	0	0	0	0	0	0
D	2	0	1	6	14	0	0	3	13	0	0	0	0	0	4	1	0	0	4	4	1	1	4	0	0	0
E	16	3	8	26	3	5	2	7	6	0	0	4	5	10	5	4	1	22	9	12	2	4	8	0	3	0
F	3	0	0	1	0	1	0	0	5	0	0	0	0	0	10	0	0	3	0	3	1	0	0	0	0	0
G	5	1	0	0	5	0	1	4	0	0	0	1	0	0	3	1	0	6	0	0	0	0	1	0	0	0
H	24	0	0	0	32	2	0	0	7	0	0	1	0	0	8	0	0	0	0	5	1	0	0	0	0	0
I	0	1	8	1	3	0	2	0	0	0	0	2	0	16	9	0	0	2	9	8	0	7	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
L	3	0	0	4	6	0	0	1	6	0	0	8	2	3	3	0	0	1	0	1	0	1	1	0	2	0
M	2	1	0	0	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
N	10	0	5	9	4	1	9	0	2	0	0	3	2	4	12	0	0	0	4	8	1	1	0	0	2	0
O	1	3	1	3	0	6	1	1	0	0	0	1	4	20	2	5	0	17	3	13	7	2	3	0	0	0
P	0	0	0	0	5	0	0	0	0	0	0	4	0	0	4	0	0	2	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
R	8	0	0	1	26	4	3	0	1	0	1	3	0	1	6	2	0	0	5	12	3	0	3	0	0	0
S	4	2	2	0	10	2	1	6	1	0	1	0	0	1	4	0	0	1	0	8	1	0	0	0	0	0
T	4	1	4	1	11	5	1	47	18	0	0	3	0	2	11	1	0	2	0	9	0	0	5	0	1	0
U	1	0	0	0	0	0	3	0	0	0	0	2	0	3	0	0	0	5	5	2	0	0	0	0	0	0
V	2	0	0	0	17	0	0	0	3	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
W	2	1	0	0	11	0	0	8	1	0	0	0	0	1	2	0	0	0	0	1	0	0	1	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	2	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIG. 1. The digram frequency matrix for Lincoln's Gettysburg address.

the frequency with which the digram $u_i u_j$ occurs in positions p, $p+1$ modulo k. All of these occurrences are encoded into the same digram $u_{c_p(i)} u_{c_{p+1}(i)}$ in the cryptogram M_c. By an argument similar to that given above for the simple substitution, if A_p is the digram frequency matrix for digrams occurring in position p, $p+1$ in M, then $C_p A_p C_{p+1}^T$ is the corresponding digram frequency matrix for the cryptogram M_c. It has the same coefficients as A_p, but its rows are permuted by C_p, and its columns by C_{p+1}. Its singular values are the same as those of A_p, and its jth left and right singular vectors are, respectively, $C_p x_{pj}$ and $C_{p+1} y_{pj}$, where x_{pj} and y_{pj} are those of A_p. If the classification scheme described above, applied to x_{pj} and $y_{(p-1)j}$ classifies u_i as a vowel, then it will also, applied to $C_p x_{pj}$ and $C_p y_{(p+1)j}$, classify its encoding $u_{c_p(i)}$ as a vowel.

If this is done for each p, then the cryptanalyst has, for each p, a classification of the letters in position p of the cryptogram, into vowels and consonants. The classification is, of course, only as reliable as it would have been if it had been applied to the original message M, or, more particularly, to those subsequences of M consisting of the digrams occurring in a particular position p, $p+1$. If these are representative samples of the digrams occurring in the language of M, then the classification is as reliable as it is when applied to equally representative plaintexts, as we have done in sections 7 and 8.

1	81.7256
2	53.5189
3	45.1604
4	31.1073
5	19.9258
6	18.8196
7	16.8777
8	12.1174
9	10.1647
10	9.4667
11	7.6957
12	6.2491
13	4.7882
14	2.7120
15	2.1048
16	1.7556
17	1.6395
18	0.9360
19	0.8129
20	0.4996
21	0.3232
22	0.2069
23	0.0148
24	0.0000
25	0.0
26	0.0

FIG. 2. The singular values.

7. Experimental Results. Fig. 1 is the digram frequency matrix for Lincoln's Gettysburg Address, a text of 1,148 characters. Notice, for example, that "th," with 47 occurrences, is the most frequent pair, that "q" occurs only once, and that "j," "x" and "z" do not occur at all.

Fig. 2 gives the singular values of the matrix in Fig. 1. Since three letters are missing, the matrix has rank 23 at most, and 3 of the singular values should be zero. The subroutine finds two exact zero values and one value, the size of roundoff error on the computer. (Exact zeros are printed as 0.0, while numbers less than 10^{-5} are printed as 0.00000.)

It certainly cannot be claimed that the singular values decrease rapidly. In fact, the rank two approximation only vaguely resembles the original matrix. Nevertheless, useful information can be obtained from the first two pairs of singular vectors.

Fig. 3 shows the first singular vectors x_1 and y_1, together with the frequency vector f. It can be seen that the components of the two singular vectors are roughly equal, and are roughly proportional to the components of the frequency vector. Thus, even though the matrix is not particularly well approximated by the first term of its SVD, the singular vectors still retain the properties predicted by the rank one theory.

A	0.3275	0.3221	102.
B	0.0470	0.0441	14.
C	0.1200	0.1135	31.
D	0.2011	0.2259	58.
E	0.4394	0.4517	165.
F	0.0875	0.1065	27.
G	0.0966	0.0799	28.
H	0.3481	0.3378	80.
I	0.1830	0.2339	68.
J	0.0000	0.0000	0.
K	0.0099	0.0097	3.
L	0.1203	0.1218	42.
M	0.0607	0.0468	13.
N	0.2165	0.2435	77.
O	0.2387	0.2563	93.
P	0.0522	0.0564	15.
Q	0.0007	0.0054	1.
R	0.2954	0.2493	79.
S	0.1683	0.1391	44.
T	0.4453	0.4367	126.
U	0.0532	0.0597	21.
V	0.1167	0.0817	24.
W	0.1210	0.1044	28.
X	0.0	0.0	0.
Y	0.0339	0.0344	10.
Z	0.0	0.0	0.

FIG. 3. The first right and left singular vectors and the letter frequency vector.

A	-0.5097	0.1574
B	0.0412	-0.0386
C	0.0719	-0.0788
D	0.1436	-0.2163
E	-0.3306	0.5305
F	-0.0006	-0.0198
G	0.0521	-0.0623
H	0.3877	0.3293
I	-0.2070	0.1791
J	0.0000	-0.0000
K	0.0033	-0.0049
L	0.0298	-0.0853
M	0.0634	-0.0613
N	-0.0649	-0.3983
O	-0.3891	0.1070
P	0.0385	-0.0535
Q	-0.0010	-0.0062
R	0.1692	-0.3402
S	0.0801	-0.1462
T	0.3785	-0.3878
U	-0.0860	-0.0516
V	0.1884	-0.1405
W	0.1559	-0.0112
X	0.0	0.0
Y	-0.0045	-0.0215
Z	0.0	0.0

FIG. 4. The second right and left singular vectors. The sign patterns identify vowels and consonants.

Fig. 4 shows the second singular vectors x_2 and y_2. The alternating sign patterns predicted by the rank two theory are clearly evident. Figs. 5 and 6 give a graphical summary of the quantitative information in Fig. 4 by plotting each letter at a point in the two-dimensional plane determined by its components in x_2 and y_2. Thus, "a" is plotted at coordinates $(-0.5097, 0.1574)$, "b" at coordinates $(0.0412, -0.0386)$, and so on. The more frequent letters are in Fig. 5 and the less frequent letters in Fig. 6. The box in both figures has corners at $(\pm 0.1, \pm 0.1)$.

Since they fall in the same quadrant (the second), the letters "a," "e," "i" and "o" should all be classified as either vowels or consonants, and we have, of course, chosen to call them vowels.

The letters "h," "n," "u" and "y" must be called neuter because the corresponding signs in the two vectors agree. The letter "q" occurred only once. The classification of q as neuter is nevertheless interesting. Its one occurrence is in the word "equal." One might expect it to be classified as a consonant, since it occurs between two vowels. Note, however, that the algorithm does not recognize u as a vowel. It is classified as neuter. This is, no doubt, attributable to the very high frequency of the digram ou, which accounts for 7 of the 21 occurrences of u. The letter "h"

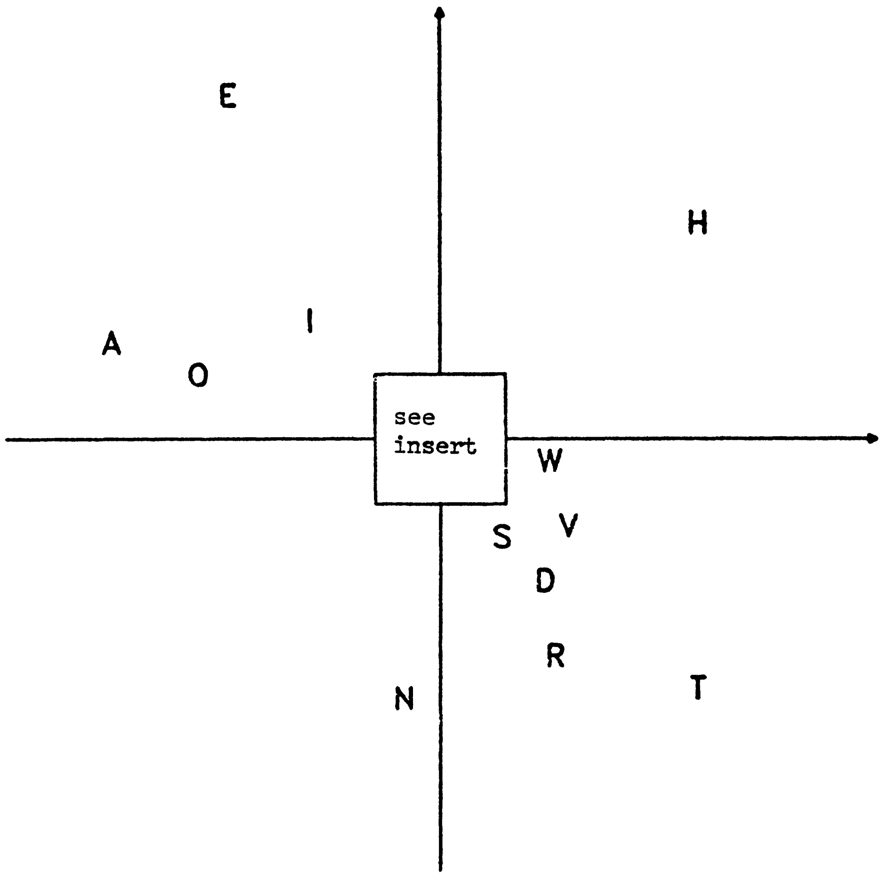


FIG. 5. The more frequent letters, plotted with coordinates from Fig. 4.

clearly shows a tendency to be followed by a vowel, and to be preceded by a consonant. The letter "n" shows a weak tendency to be followed by a consonant and a strong tendency to be preceded by a vowel. The other three neuter letters occur infrequently. The letter "j," "x," and "z" are seen not to occur at all. The remaining 14 letters are classified as consonants.

If a simple substitution cryptogram were made from text such as the Gettysburg Address, the digram matrix A would be replaced by PAP^T for some unknown permutation matrix P. As shown in Section 6, the singular vectors of the transformed matrix would classify the encoding of each letter in the same way as x_2 and y_2 classified the letters of the plaintext.

8. Other Languages. The same experiment was performed on texts of approximately one thousand characters each, written in five other languages selected for their diversity. The classifications of letters resulting from these tests are tabulated below.

Language	*Vowels*	*Consonants*	*Neuter*	*Absent*
Hawaiian	AEIOU	HKLMNPW		BCDFGJQ RSTVXYZ
Japanese	AEIOU	BDGHKMNRSTWYZ		CFJLPQVX
German	AEFOPU	BHIKLNRV	CDGJMSTWZ	QXY
Spanish	AEO	CDFGLMNQRSXYZ	BHIJPTUV	KW
Finnish	AEIOUY	DHJKLMNPRSTV	B	CFGQWXZ

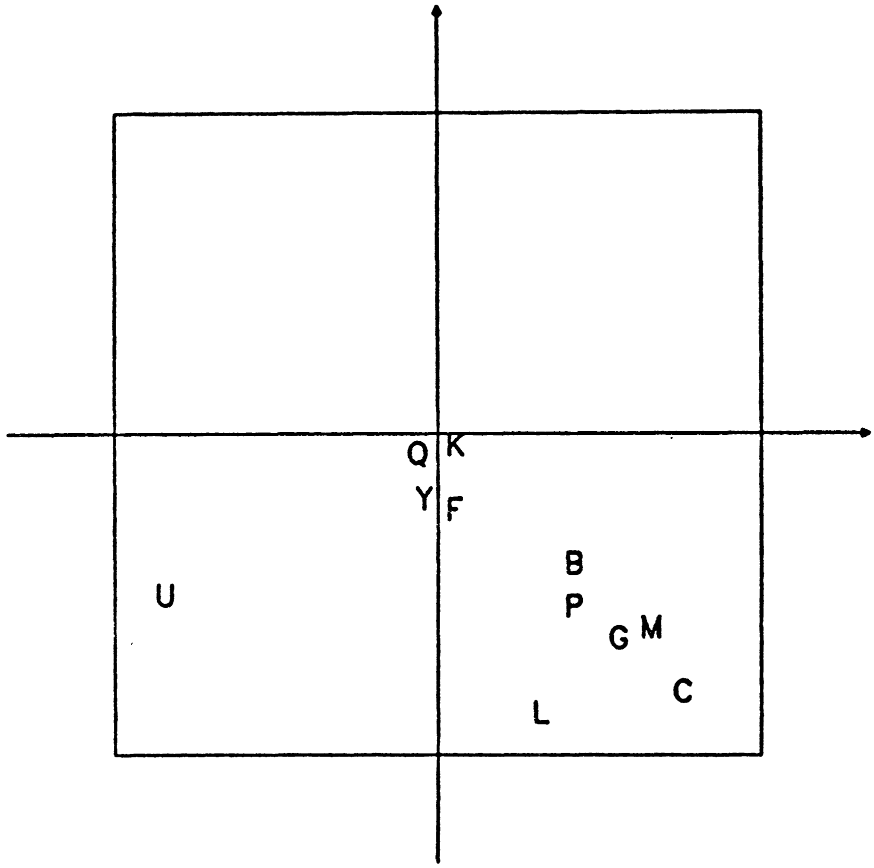


FIG. 6. The less frequent letters. The scale is 5 times that of Fig. 5.

None of the discrepancies is very surprising. Several of them are associated with letters which occurred so infrequently that no significance can be attached to their classification. In the Finnish example, B occurred only once, at the beginning of a word of foreign origin. It followed a final N in the preceding word and preceded an E. In the German text, P occurred only once, and that occurrence was in the very Germanic trigram SPR. Occurring only between two consonants, it is not surprising that the analysis classified it as a vowel. The most common neighbor of F, on both sides, in the selected sample, is R. The classification of I is confused because of the very high frequency of the diagrams IE and EI. More generally, the German language does not share the aversion of the other languages to consecutive consonants, and consequently many German letters fall into the neuter category. The classification of Y as a vowel in the Finnish sample is not surprising; Y has a value in Finnish that is hardly distinguishable to a non-Finnish ear, from that of U. The neuter classification of I and U in the Spanish example is clearly attributable to the high frequency of vowel-vowel digrams in which U or I is the first letter, having a value equivalent to W or Y in English. The perfect performance of the algorithm in Japanese and Hawaiian is clearly the result of their rigorously observed exclusion of consonant-consonant digrams. Most of the Japanese Hiragana characters are transliterated as a consonant-vowel digram.

9. Conclusions. The second singular vectors in the singular value decomposition of the digram frequency matrix provide the cryptanalyst with a helpful and surprisingly reliable way to classify the letters in a cryptogram as vowels or consonants, if the encoding algorithm is simple

substitution or k-alphabetic substitution, with k known, and if the text if vfc text. Texts written in many natural languages, with certain exceptions, tend to be vfc texts. Near-exceptions are German (and, presumably, other germanic languages) in which the vfc character is somewhat diminished by the frequency of consonant-consonant digrams, and Spanish (and, presumably, other romance languages) in which certain vowel-vowel pairs are frequent. Despite these deviations from the vfc rule, the second singular vectors classify correctly most of the letters which occur with a frequency high enough to be statistically significant.

A computer program which uses the SVD as the starting point in an automated, heuristic approach to solving cryptograms is described by Schatz [7].

References

1. H. C. Andrews and C. L. Patterson, Outer product expansions and their uses in digital image processing, this MONTHLY, 82 (1975) 1–12.
2. H. F. Gaines, Cryptanalysis, Dover, New York, 1956.
3. G. E. Forsythe, M. A. Malcolm, and C. B. Moler, Computer Methods for Mathematical Computations, Prentice-Hall, Englewood Cliffs, NJ, 1977.
4. Richard S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962.
5. A. Sinkov, Elementary Cryptanalysis, A Mathematical Approach, New Mathematical Library, vol. 22, Mathematical Association of America, Washington, D.C., 1968.
6. J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, LINPACK Users' Guide, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.
7. Bruce R. Schatz, Automated analysis of cryptograms, Cryptologia, 1 (1977) 116–142.

WHAT IS THE GEOMETRY OF A SURFACE?

ROGER FENN
School of Mathematics and Physical Sciences, University of Sussex, Falmer, Brighton BN1 9QH, England

1. In this article we shall assume the definition of a geometry proposed by Klein in his Erlangen program, that is, a geometry is a group of transformations of space together with the propositions left invariant by this group. The space in question will be the two-dimensional plane. By a surface we shall mean a closed compact orientable surface of genus γ. That is a sphere with γ handles attached. The fact that two such surfaces are homeomorphic if and only if their genera are equal was probably known to Riemann. A modern proof can be found in Massey's book [**5**, Chapter 1].

The exact result which will be proved is the following:

THEOREM 1. *Let $\gamma > 1$. Then there is a group of hyperbolic translations of the hyperbolic plane such that the space of orbits under these translations is homeomorphic to a surface of genus γ.*

Moreover, there is a compact polygon in the plane whose translates under the group are distinct for distinct group elements and which form a network of nonoverlapping polygons covering the whole hyperbolic plane.

Our aim is to prove the above by means as elementary as possible, in particular without using deep results from the theory of functions.

In order to illustrate the general result, we consider firstly the case where $\gamma = 1$. Here the

---

Roger Fenn is a Lecturer in Mathematics at the University of Sussex, England. He received his Ph.D. in 1968 under the direction of John Reeve and is presently writing a book on Geometric Topology. He believes people should be at the same time renaissance polymaths and twentieth century specialists insofar as this is possible. He is married, with three children, and enjoys playing chess, singing madrigals, and walking.

substitution or k-alphabetic substitution, with k known, and if the text if vfc text. Texts written in many natural languages, with certain exceptions, tend to be vfc texts. Near-exceptions are German (and, presumably, other germanic languages) in which the vfc character is somewhat diminished by the frequency of consonant-consonant digrams, and Spanish (and, presumably, other romance languages) in which certain vowel-vowel pairs are frequent. Despite these deviations from the vfc rule, the second singular vectors classify correctly most of the letters which occur with a frequency high enough to be statistically significant.

A computer program which uses the SVD as the starting point in an automated, heuristic approach to solving cryptograms is described by Schatz [7].

References

1. H. C. Andrews and C. L. Patterson, Outer product expansions and their uses in digital image processing, this MONTHLY, 82 (1975) 1–12.

2. H. F. Gaines, Cryptanalysis, Dover, New York, 1956.

3. G. E. Forsythe, M. A. Malcolm, and C. B. Moler, Computer Methods for Mathematical Computations, Prentice-Hall, Englewood Cliffs, NJ, 1977.

4. Richard S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962.

5. A. Sinkov, Elementary Cryptanalysis, A Mathematical Approach, New Mathematical Library, vol. 22, Mathematical Association of America, Washington, D.C., 1968.

6. J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, LINPACK Users' Guide, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.

7. Bruce R. Schatz, Automated analysis of cryptograms, Cryptologia, 1 (1977) 116–142.

WHAT IS THE GEOMETRY OF A SURFACE?

ROGER FENN

School of Mathematics and Physical Sciences, University of Sussex, Falmer, Brighton BN1 9QH, England

1. In this article we shall assume the definition of a geometry proposed by Klein in his Erlangen program, that is, a geometry is a group of transformations of space together with the propositions left invariant by this group. The space in question will be the two-dimensional plane. By a surface we shall mean a closed compact orientable surface of genus γ. That is a sphere with γ handles attached. The fact that two such surfaces are homeomorphic if and only if their genera are equal was probably known to Riemann. A modern proof can be found in Massey's book [5, Chapter 1].

The exact result which will be proved is the following:

THEOREM 1. *Let $\gamma > 1$. Then there is a group of hyperbolic translations of the hyperbolic plane such that the space of orbits under these translations is homeomorphic to a surface of genus γ.*

Moreover, there is a compact polygon in the plane whose translates under the group are distinct for distinct group elements and which form a network of nonoverlapping polygons covering the whole hyperbolic plane.

Our aim is to prove the above by means as elementary as possible, in particular without using deep results from the theory of functions.

In order to illustrate the general result, we consider firstly the case where $\gamma = 1$. Here the

---

Roger Fenn is a Lecturer in Mathematics at the University of Sussex, England. He received his Ph.D. in 1968 under the direction of John Reeve and is presently writing a book on Geometric Topology. He believes people should be at the same time renaissance polymaths and twentieth century specialists insofar as this is possible. He is married, with three children, and enjoys playing chess, singing madrigals, and walking.

geometry is Euclidean and the surface is a torus.

In the Euclidean plane let a and b be the translations given in coordinate form by $a(x, y) = (x + 1, y)$ and $b(x, y) = (x, y + 1)$. Then a and b generate a group T called the torus group. This is the free Abelian group with basis $\{a, b\}$. Let S_T be the quotient space under the action of T. That is (x, y) and (x', y') are to be identified in S_T if and only if $(x', y') = g(x, y)$ for some $g \in T$. Equivalently (x, y) and (x', y') are identified if and only if $x - x'$ and $y - y'$ are integers. Then S_T is obtained from the unit square Q, $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ by identifying according to the rule indicated in Figure 1.

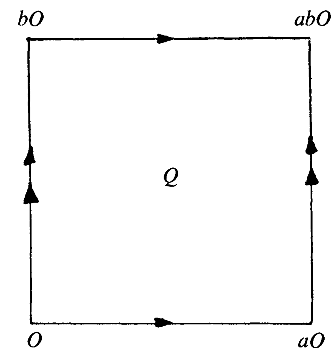


FIG. 1. The fundamental region for the torus group.

So S_T is a torus.

In technical language the *universal cover* of the torus is the Euclidean plane and the group of *covering translations* corresponding to the *fundamental group* is generated by the Euclidean isometries a and b. The square in Figure 1 is called a *fundamental region*. The translations of the fundamental region by the covering translations cover the Euclidean plane in a network of

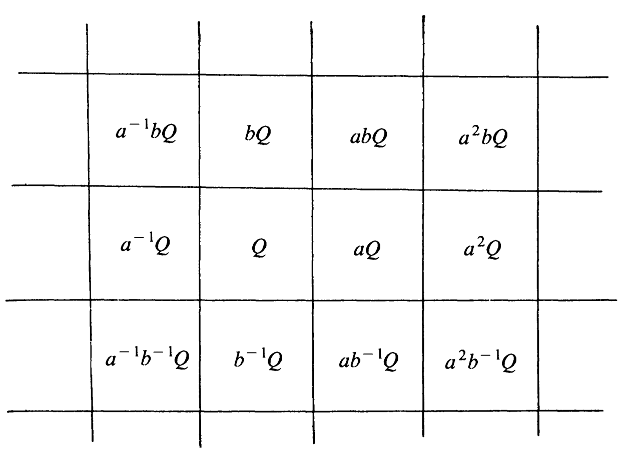


FIG. 2. The translates of the fundamental region by the torus group.

nonoverlapping squares all of the same size: see Figure 2.

If $\gamma = 0$, then the surface is a sphere and the geometry is *spherical* or *elliptic*. We shall not consider this case here. The interested reader can read the books by Coxeter or Guggenheimer [**2**], [**3**].

The idea in the sequel which was known to Poincaré, Klein and Fricke is to generalize these notions so that the torus is replaced by a surface of genus $\gamma > 1$ and the Euclidean plane is replaced by the hyperbolic plane. The fundamental region will become a hyperbolic polygon with 4γ sides and the Euclidean translations will become hyperbolic translations which cover the hyperbolic plane with an infinite network of copies of the fundamental polygon.

2. A Quick Review of Hyperbolic Geometry. In this section a brief description of the hyperbolic plane will be presented. By and large, proofs will be omitted. There are many excellent books on hyperbolic geometry to which the interested reader may refer for more details. A suitable reference might be Coxeter [**2**] or the appropriate chapter of Caratheodory [**1**, 80–88].

The model used here is due to Poincaré. Let $\mathcal{H}$ be the interior of the unit disc in the complex plane $\mathbf{C}$. The points of $\mathcal{H}$ are the *points* of the hyperbolic plane. The points on the unit circle S^1 are often called the *points at infinity*. The *hyperbolic lines* or $\mathcal{H}$-lines for short are the arcs in $\mathcal{H}$ of circles orthogonal to S^1. Also included as $\mathcal{H}$-lines are diameters of $\mathcal{H}$.

Two points A, B of $\mathcal{H}$ determine a unique $\mathcal{H}$-line. Write AB for the $\mathcal{H}$-arc joining them. This notation will also be used if one or both of A, B lie at infinity. If two $\mathcal{H}$-lines meet at A inside $\mathcal{H}$, their circle extensions also meet at the inverse point $1/\bar{A}$ outside $\mathcal{H}$.

If two $\mathcal{H}$-lines meet only at infinity, they are called parallel. If two $\mathcal{H}$-lines fail to meet, they are sometimes called *ultraparallel*.

If A, B, C, D are four points of $\mathbf{C}$, their *cross ratio* is defined to be

$$\delta(A, B, C, D) = (A - C)(B - D)/(A - D)(B - C).$$

Let A, B lie in $\mathcal{H}$. The line through A and B meets the circle at infinity in two more points C, D: see Figure 3.

FIG. 3.

The cross ratio of these points, lying on a circle, is real and positive. Define the *hyperbolic distance* between A and B by

$$d(A, B) = \log \delta(A, B, C, D).$$

The angle between two $\mathcal{H}$-lines is measured in this model in exactly the same way as the ordinary Euclidean angle. It is an important fact in the history of noneuclidean geometry that the three angles of a triangle have sum less than π. This is a consequence of the Gauss-Bonnet theorem and the fact that the hyperbolic plane has constant negative curvature. It is also useful to consider *ideal triangles* in which one or more vertices are at infinity. In this manner polygons with angles equal to zero can be considered.

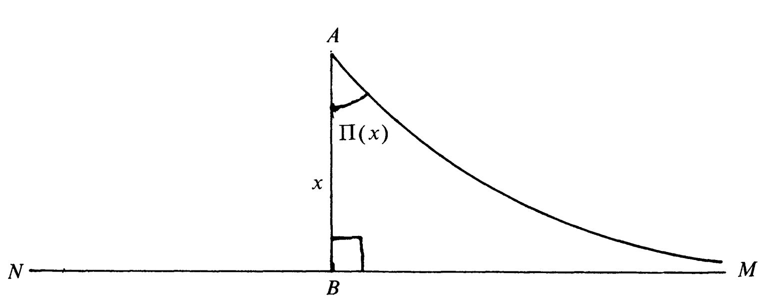


FIG. 4. The angle $\Pi(x)$.

Let NM be an $\mathcal{H}$-line where NM are at infinity and let A be a point of $\mathcal{H}$ not on the line NM. Let B be the foot of the perpendicular from A to NM. If AM is one of the $\mathcal{H}$-lines through A parallel to NM and x is the distance AB, let $\Pi(x)$ denote the angle BAM: see Figure 4. The function $\Pi(x)$ is called the *angle of parallelism*. For Euclidean geometry this would always be $\pi/2$. Analytically $\Pi(x) = 2\tan^{-1}(e^{-x})$.

Transformations of the hyperbolic plane which preserve orientation and arc length are called (orientation preserving) isometries. They will also preserve area, angles and the sense of angles. All such isometries have the analytical form

$$Z \to W = e^{i\phi}(A - Z)/(1 - \bar{A}Z), \quad |A| < 1$$

and hence can be thought of as acting on the whole complex plane $\mathbb{C}$. They constitute the group of bilinear transformations

$$W = (\alpha Z + \beta)/(\gamma Z + \delta), \quad \alpha\delta - \beta\gamma \neq 0$$

which preserve the unit disc, $|Z| \leqslant 1$ and hence its interior $\mathcal{H}$. There is a unique bilinear transformation taking the points A, B, C to P, Q, R given by

$$(W - Q)(P - R)/(W - R)(P - Q) = (Z - B)(A - C)/(Z - C)(A - B).$$

The analytic form expresses the fact that the cross ratio $\delta(Z, A, B, C)$ is preserved. The uniqueness follows from the fact that the only bilinear transformation with three fixed points is the identity.

Let A, B and A', B' be two pairs of points with $d(A, B) = d(A', B')$. Let C, D and C', D' be the corresponding points at infinity as in Figure 3. Then the unique bilinear transformation which takes A, B, C into A', B', C' preserves cross ratios and so takes D into D'. This isometry takes the circle through D, C perpendicular to $DABC$ into the circle through D', C' perpendicular to $D'A'B'C'$. But both circles are the same, namely the circle at infinity. Consequently, there is a unique isometry of $\mathcal{H}$ taking two given points to two other given points the same distance apart.

The *discriminant* Δ of an isometry is given by

$$\Delta = \sin^2(\phi/2)/(1 - A\bar{A}).$$

The value of Δ divides the set of isometries into three classes.

Elliptic isometries, or rotations, $\Delta < 1$

These have two fixed points ρ and $1/\bar{\rho}$ separated by the unit circle. An elliptic isometry can be thought of as a rotation about the fixed point lying inside $\mathcal{H}$ through an angle equal to $2\cos^{-1}\sqrt{\Delta}$. The set of orbits is the pencil of circles (and one line) with degenerate circles ρ and $1/\bar{\rho}$. The one line is the perpendicular bisector of the two fixed points: see Figure 5.

The orthogonal pencil to the orbits is permuted by the isometry.

The elliptic isometry with fixed point A and angle of rotation ϕ is given by

$$(W - A)/(W - 1/\bar{A}) = e^{i\phi}(Z - A)/(Z - 1/\bar{A}).$$

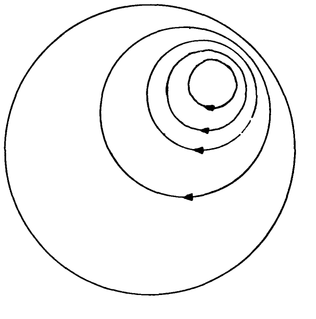


FIG. 5. Orbits of an elliptic isometry.

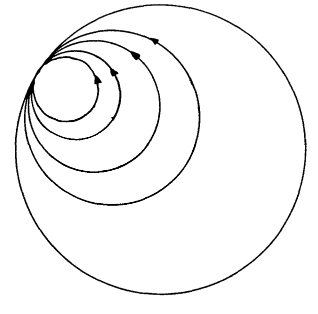


FIG. 6. Orbits of a parabolic isometry.

Parabolic isometries, $\Delta = 1$

These have a unique fixed point lying on the unit circle. The orbit set is the pencil of circles tangent to S^1 at the fixed point. Circles touching the circle at infinity are called *horocycles*. Again the orthogonal family is permuted by the isometry: see Figure 6.

A parabolic isometry is the "limit" of an elliptic isometry as the internal fixed point tends to infinity.

A parabolic isometry with fixed point A is given by

$$iA/(W-A) = iA/(Z-A) - k$$

where k is an arbitrary nonzero real number.

Hyperbolic isometries or translations, $\Delta > 1$

This case is dual to the elliptic case. Hyperbolic isometries have two fixed points lying on the unit circle. The orbit set is the pencil of circles (and the one straight line) passing through these fixed points. One of these circles defines an $\mathcal{H}$-line called the *axis* of the isometry: see Figure 7. The other circles are sometimes called *hypercycles* or equidistant curves. They are the set of points "equidistant" from the axis.

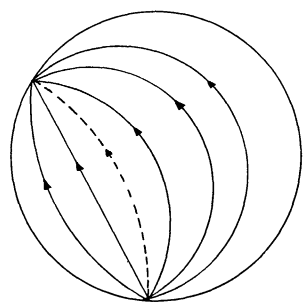


FIG. 7. Orbits of a hyperbolic isometry. The axis is dotted.

A hyperbolic isometry can be thought of as a translation along the axis through a distance 2 $\cosh^{-1}\sqrt{\Delta}$.

One fixed point of a hyperbolic isometry is a source and the other a sink. This allows the following useful characterization of hyperbolic isometries: let L_y be the portion of the y axis in $\mathcal{H}$ oriented upwards in the usual manner. Then L_y divides $\mathcal{H}$ into two regions, a right and a left. In general any hyperbolic line divides $\mathcal{H}$ into two regions. Let h be any isometry which takes L into L_y and preserves the orientation of L and L_y. Call the region divided by L *right* if h takes it into the right side of L_y. Similarly for left.

Suppose now that an isometry h takes the oriented line L_1 into the oriented line L_2 preserving their orientation. The lines L_1 and L_2 are not to meet even at infinity, i.e., they are ultraparallel. Let H_1 be the part of $\mathcal{H}$ to the right of L_1. Define H_2 similarly. If $H_1 \subset H_2$ or $H_2 \subset H_1$, then h is hyperbolic. In the first case the source lies in $\mathcal{H}_1$, in the second case the source lies in $\mathcal{H}_2$.

Another useful characterization is that for all other types of isometry h, there are points Z and hZ arbitrarily close to one another.

A hyperbolic isometry with fixed points A and B is given by

$$(W-A)/(W-B) = k(Z-A)/(Z-B)$$

for some real $k \neq 0$.

3. The Construction of the Fundamental Polygon P. To help understand the following the reader should consult Figure 8, which illustrates the case $n = 8$. Let $X_1, X_2, \ldots, X_n$ be the nth roots of unity where $n = 4\gamma$, $\gamma > 1$. Let $Y_1, Y_2, \ldots, Y_n$ be points on S^1 so that Y_i is midway

between X_i and X_{i+1}. (Here i is a cyclic index taking the values $i = 1, 2, \ldots, n$, mod n.) Consider the $\mathcal{H}$-lines X_iX_{i+1}. These form a regular n-gon with vertices at infinity and internal angle equal to zero. The half spaces of the $\mathcal{H}$-lines Y_iY_{i+2} determine a regular compact n-gon with internal angle equal to $2\alpha'$, say. Let the vertices of this polygon be $Z_1', \cdots Z_n'$. In the diagram Z_2' is B and Z_3' is A. Let Z_i be points on the lines OX_i, equidistant from O and lying between Z_i' and X_i', $i = 1, \ldots, n$. Define P to be the polygon with vertices $Z_1, \ldots, Z_n$. Let 2α be the internal angle of P. The triangle OZ_3Z_2 has larger area than the triangle OAB so that $\alpha < \alpha'$. By continuity 2α can take any value in the range $0 \leqslant 2\alpha < 2\alpha'$.

By considering the triangle OAB it can be seen that

(1) $$2\alpha' + 2\pi/n < \pi.$$

By considering the triangle AY_2B it can be seen that

(2) $$2\pi - 4\alpha' < \pi.$$

The inequalities 1 and 2 imply

$$2\pi/4 < 2\alpha' < \pi(1 - 2/n).$$

By continuity P may be chosen so that its internal angle $2\alpha = 2\pi/n$.

4. The Hyperbolic Isometries a_i and b_i. Let the vertices of P be $Z_1, Z_2, \ldots, Z_n$ in anti-clockwise cyclic order. Let a_i be the isometry which takes $Z_{4i-1}Z_{4i}$ to $Z_{4i-2}Z_{4i-3}$ and let b_i take

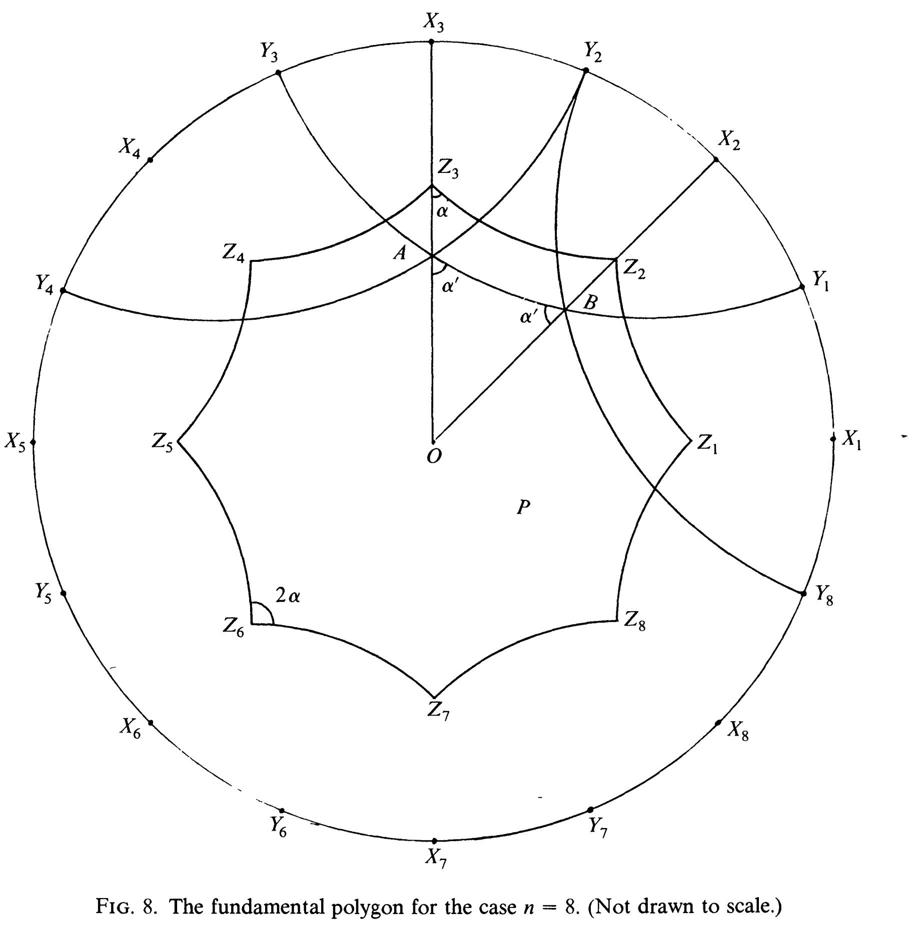


FIG. 8. The fundamental polygon for the case $n = 8$. (Not drawn to scale.)

$Z_{4i-2}Z_{4i-1}$ to $Z_{4i}Z_{4i+1}$, $i = 1, \ldots, \gamma$. For an illustration of a_1 and b_1 see Figure 9.

Looking at Figure 8 it can be seen that the half spaces defined by the $\mathcal{H}$-lines through Z_3Z_4 and through Z_2Z_1 are nested. Therefore a_1 (and similarly b_1) is hyperbolic. By symmetry all a_i and b_i are hyperbolic. However, this will also follow later as part of a more general result.

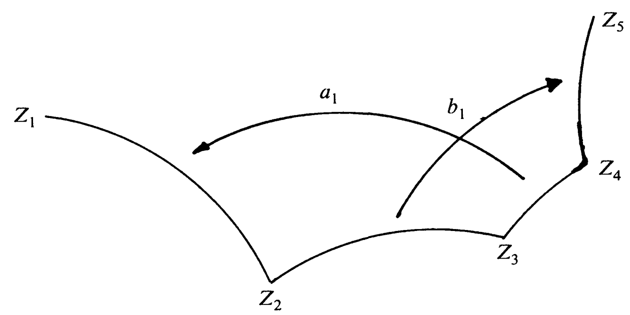


FIG. 9. The isometries a and b.

Let

$$c_i = a_i b_i a_i^{-1} b_i^{-1} = [a_i, b_i], \quad i = 1, \ldots, \gamma.$$

Then $c_i Z_{4i+1} = Z_{4i-3}$.

Consider the group G of all isometries generated by $a_1, b_1, a_2, b_2, \ldots, a_\gamma, b_\gamma$. If g is an element of G, let $gP = g(P)$ be the image of P under the isometry g.

As g varies over the 4γ elements

$$g = a_1, a_1b_1, a_1b_1a_1^{-1}, \ldots, c_1c_2 \cdots c_{\gamma-1}a_\gamma b_\gamma a_\gamma^{-1}, c_1c_2 \cdots c_{\gamma-1}c_\gamma = 1,$$

the polygons gP share the vertex Z_1 and the corresponding angles are lined up clockwise around Z_1: see Figure 10 for the case $\gamma = 2$. The interiors of adjacent polygons are disjoint for the following reasons: Consider say the intersection of a_1b_1P and $a_1b_1a_1^{-1}P$. This is a translation of the intersection of P and $a_1^{-1}P$ which only consists of their common edge. Since P was chosen to have internal angle equal to $2\pi/4\gamma$, the 4γ polygons gp form a full neighbourhood of Z_1. So in

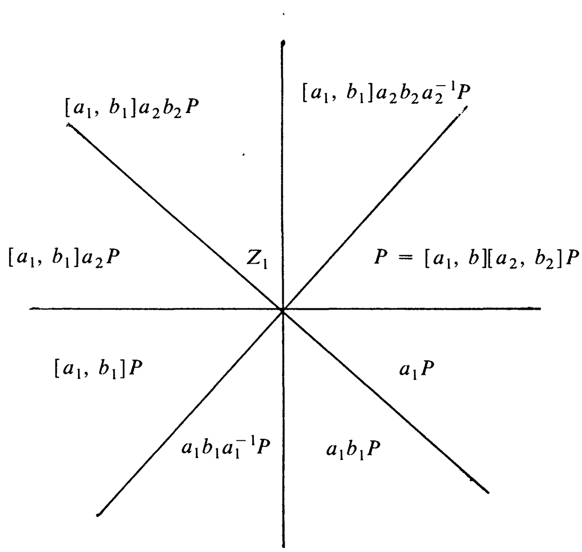


FIG. 10. The polygons all fit round a vertex without overlapping.

particular if $g = c_1 c_2 \cdots c_\gamma$, then $gP = P$. A quick calculation shows that $gZ_1 = Z_1$. Since g is an orientation preserving isometry, it follows that $gZ_i = Z_i$ for all vertices Z_i of P. But the only isometry of $\mathcal{H}$ with more than two fixed points is the identity. So $g = c_1 c_2 \cdots c_\gamma = 1$.

By a cyclic permutation of the 4γ factors in $c_1 c_2 \cdots c_\gamma$ the same is true for every other vertex of P. Furthermore if gZ_i is a vertex of gP, the polygons which form a neighbourhood of Z_i are transformed to form a neighbourhood of gZ_i.

5. The Tiling of the Plane by the Polygons gP. In this section it will be shown that the translates gP of the fundamental polygon P cover the whole plane without overlapping. This tiling is constructed implicitly in Siegel's book [**9**, Chapter 12]. See also the book by Magnus for $\gamma = 2$ [**4**, p. 187]. It has already been shown that the polygons fit nicely along the edges and about the vertices. It remains to show that the polygons do not overlap globally. For the case $\gamma = 2$ see Figure 11. The idea behind the proof is to consider the space $\tilde{\mathcal{H}}$ obtained from the union of the polygons by not allowing them to overlap at a distance and then to show that $\tilde{\mathcal{H}}$ actually is $\mathcal{H}$!

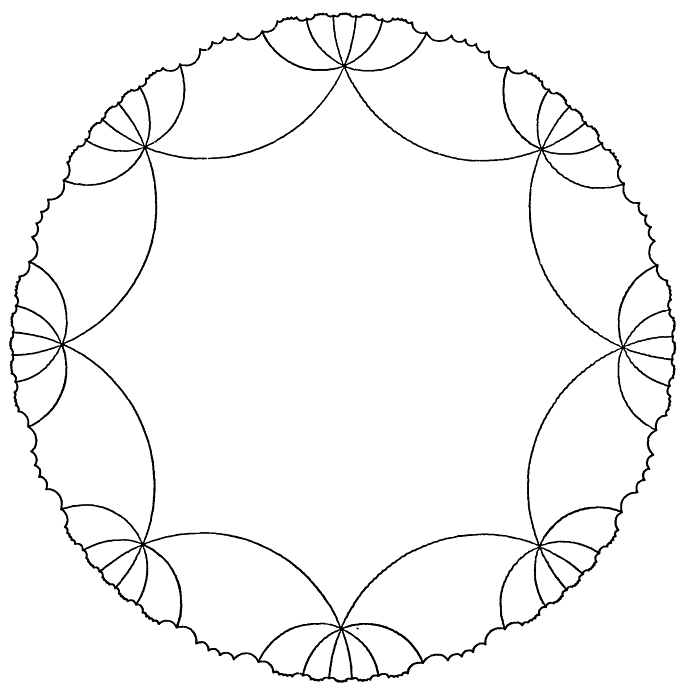


FIG. 11. The network of polygons when $\gamma = 2$.

More specifically let $g\tilde{P}$ be a copy of the polygon gP for each g in G. Assume initially that the gP are mutually disjoint. Let X be the set of generators of G and let X^{-1} be the set of inverses. So

$$X = \{a_1, a_2, \ldots, a_\gamma, b_1, b_2, \ldots, b_\gamma\},$$

$$X^{-1} = \{a_1^{-1}, a_2^{-1}, \ldots, a_\gamma^{-1}, b_1^{-1}, b_2^{-1}, \ldots, b_\gamma^{-1}\}.$$

Then $g\tilde{P}$ and $h\tilde{P}$ are to be glued along an edge if and only if $g^{-1}h \in X \cup X^{-1}$. In this case let e be the edge of P in common with $g^{-1}hP$. Then gP and hP will have the edge ge in common: Glue $g\tilde{P}$ and $h\tilde{P}$ together using their copy of this edge.

Let $\tilde{\mathcal{H}}$ denote the resulting space. So $\tilde{\mathcal{H}} = \cup g\tilde{P}$ with the above identifications. Notice that $\tilde{\mathcal{H}}$ has the same local structure as $\mathcal{H}$. There is a map $p\colon \tilde{\mathcal{H}} \to \mathcal{H}$ which takes $\tilde{P}$ identically to its copy P. Locally distances, angles and indeed geometry are identical and are identified by the map p.

The polygons $g\tilde{P}$ cover $\tilde{\mathcal{H}}$ completely without overlapping.

Firstly let us show that p is onto. Because $\mathcal{H}$ is connected, it is only necessary to show that $\cup gP$ is both open and closed in $\mathcal{H}$. It is not hard to see that $\cup gP$ is open. This is because if the point B lies in gP, it either lies in its interior, or in the interior of an edge of gP which is also the edge of an adjacent hP, $gh^{-1} \in X \cup X^{-1}$, or it is a vertex of gP in which case it is surrounded by an open neighborhood of translates of P. So let A be a point of $\mathcal{H}$ which does not lie in $\cup gP$, but does lie in its closure $\overline{\cup gP}$. It will be shown that points such as these do not exist.

Every neighbourhood of A must meet at least one gP, in fact an infinite number of them. Let us suppose that this neighbourhood is the disc of radius ε and centre A where ε is to be a suitably chosen small number.

Now there is a positive number δ such that if B lies in gP and C is a distance less than δ from B, then either C lies in gP or C lies in hP where $gh^{-1} \in X \cup X^{-1}$. A suitable choice for δ could be one half the length of any side of P. But this leads to a contradiction if ε is put equal to δ because A was supposed not to lie in $\cup gP$.

Therefore the closure of $\cup gP$ is the whole of the hyperbolic plane and so the map p is onto $\mathcal{H}$.

It will now be shown that p has the *path lifting property* and hence is one-to-one. To have the path lifting property means that if α: $I \to \mathcal{H}$ is a path where $I = [0,1]$ is the closed interval $0 \leqslant x \leqslant 1$ and $\tilde{A} \in \tilde{\mathcal{H}}$ is a point such that $p\tilde{A} = \alpha(0)$, then there is a unique path $\tilde{\alpha}$: $I \to \tilde{\mathcal{H}}$, the lift of α such that $p\tilde{\alpha} = \alpha$ and $\tilde{\alpha}(0) = \tilde{A}$.

Let $\tilde{V}$ be a vertex of $g\tilde{P}$ in $\tilde{\mathcal{H}}$ and let $U(\tilde{V})$ be the interior of the union of all polygons with vertex $\tilde{V}$. Then the open sets $U(\tilde{V})$ as $\tilde{V}$ varies over all vertices cover $\tilde{\mathcal{H}}$. Moreover $pU(\tilde{V})$ is an exact copy of $U(\tilde{V})$ in $\mathcal{H}$ and the union of the $pU(\tilde{V})$ covers $\mathcal{H}$.

If the path α lies entirely in some $pU(\tilde{V})$ where $\tilde{x} \in U(\tilde{V})$, then the lift $\tilde{\alpha}$ is uniquely defined by

$$\tilde{\alpha} = \left(p|U(\tilde{V})\right)^{-1}\alpha.$$

If not, then by a standard compactness argument α can be broken up into a sequence of paths each of which does lie in some $pU(\tilde{V})$. That is there exists a subdivision

$$0 = t_0 < t_1 < \cdots < t_n = 1$$

of the interval [0, 1] such that the image of

$$\alpha_i = \alpha|[t_i, t_{i+1}]$$

lies in some $pU(\tilde{v})$, $i = 0,\dots, n-1$. We proceed to lift α step by step. Suppose α has been lifted for $0 \leqslant t \leqslant t_i$. Consider α_i: this has a unique lift $\tilde{\alpha}_i(t_i) = \tilde{\alpha}_{i-1}(t_i)$ because α_t lies in some $pU(\tilde{v})$. So the lift of $\alpha|[0, t_i]$ can be extended to $\alpha|[0, t_{i+1}]$. After n steps, the whole of α will be lifted.

To see the uniqueness of $\tilde{\alpha}$, consider another lift $\tilde{\alpha}'$ with $\tilde{\alpha}(0) = \tilde{\alpha}'(0)$. Let

$$\xi = \sup\{t\,|\,0 \leqslant t \leqslant 1, \tilde{\alpha}|[0,t] = \tilde{\alpha}'|[0,t]\}.$$

If $\xi = 1$, we are done. If $\xi < 1$, then for some small ε, $\alpha(\xi - \varepsilon, \xi + \varepsilon)$ lies in a $pU(\tilde{v})$ and so $\tilde{\alpha}$ and $\tilde{\alpha}'$ have a unique extension to $[1, \xi + \varepsilon)$ which is a contradiction.

Actually rather more has been shown here because clearly a continuous variation in α will induce a continuous variation in $\tilde{\alpha}$.

Now suppose that $p(\tilde{A}) = p(\tilde{B}) = A$. Join $\tilde{A}$ and $\tilde{B}$ by a path $\tilde{\alpha}$. This is the unique lift of a loop α such that $\alpha(0) = \alpha(1) = A$. Consider the 1-parameter family of loops α_t, $0 \leqslant t \leqslant 1$ given by

$$\alpha_t(x) = t\alpha(x) + (1-t)A, 0 \leqslant x \leqslant 1.$$

Here we are using the affine structure of $\mathcal{H}$ as a convex subset of $\mathbb{C}$. In this way it is possible to shrink the loop α based at A through a family of loops based at A to the trivial loop. Consider the lift $\tilde{\alpha}_t$. This starts out as a path joining $\tilde{A}$ to $\tilde{B}$ and varies continuously joining $\tilde{A}$ to $\tilde{B}$ all the time until it is a trivial path. Therefore $\tilde{A} = \tilde{B}$ and p is a one-to-one map. But we have already seen that

p is onto and so $\tilde{\mathcal{H}}$ is an exact copy of $\mathcal{H}$. In particular the polygons gP cover $\mathcal{H}$ completely without overlapping.

6. The Structure of the Group *G*. We have seen that G is in one-to-one correspondence with the polygons gP and that it has the hyperbolic isometries $a_1,\ldots, a_\gamma, b_1,\ldots, b_\gamma$ as generators, and that these satisfy the relation

$$[a_1, b_1][a_2, b_2] \cdots [a_\gamma, b_\gamma] = 1.$$

We shall now show that in a technical sense to be defined below, this is the only relation satisfied by G.

Let $\hat{G}$ be the free nonabelian group on the elements $\hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2,\ldots, \hat{a}_\gamma, \hat{b}_\gamma$ and let $\phi\colon \hat{G} \to G$ be the homomorphism which doffs hats, so

$$\phi(\hat{a}_1) = a_1, \quad \phi(\hat{b}_1) = b_1, \quad \text{etc.}$$

The element

$$\hat{c} = [\hat{a}_1, \hat{b}_1][\hat{a}_2, \hat{b}_2] \cdots [\hat{a}_\gamma, \hat{b}_\gamma]$$

lies in the kernel K of ϕ

$$K = \{\hat{g} \in \hat{G} | \phi(\hat{g}) = 1\}.$$

Let $\hat{g}$ be an arbitrary element of $\hat{G}$. Then

$$\hat{g} = \hat{x}_1 \cdots \hat{x}_k, \text{ where } \hat{x}_i \in \hat{X} \cup \hat{X}^{-1},$$
$$\hat{X} = \{\hat{a}_1, \hat{b}_1,\ldots, \hat{a}_\gamma, \hat{b}_\gamma\}, \hat{X}^{-1} = \{\hat{a}_1^{-1}, \hat{b}_1^{-1},\ldots, \hat{a}_\gamma^{-1}, \hat{b}_\gamma^{-1}\}.$$

If $\hat{g}$ lies in K, then there is a sequence of polygons

$$P_1, P_2 = x_k P_1, P_3 = x_{k-1}x_k P_1,\ldots, P_k = x_2 \cdots x_{k-1}x_k P_1$$

such that P_i and P_{i+1} have an edge in common and $P_{k+1} = x_1 P_k = P_1$. Let N be the union of the interiors of the P_i together with the union of the interiors of their common edges. Then $\mathcal{H} - N$ is divided into an unbounded outside region and a closed interior set R by the Jordan curve theorem.

Suppose firstly that the inside R consists of just one point gZ_1. Then, perhaps after some cancellation, $\hat{g}$ is a cyclic permutation of some power of the product

$$(\hat{g}\hat{a}_1\hat{g}^{-1})(\hat{g}\hat{b}_1\hat{g}^{-1}) \cdots (\hat{g}\hat{b}_\gamma^{-1}\hat{g}^{-1}) = \hat{g}\hat{c}\hat{g}^{-1}$$

or

$$(\hat{g}\hat{b}_\gamma\hat{g}^{-1}\hat{g}\hat{a}_\gamma^{-1}\hat{g}^{-1}) \cdots (\hat{g}\hat{a}_1^{-1}\hat{g}^{-1}) = \hat{g}\hat{c}^{-1}\hat{g}^{-1}$$

depending on whether the sequence of polygons is clockwise or not.

We shall now see that the general case is a combination of this special case.

Let H be the subgroup of $\hat{G}$ consisting of all products

$$\prod_{i=1}^{m} \hat{g}_i \hat{c}^{\varepsilon_i} \hat{g}_i^{-1}$$

for some $\hat{g}_i$ in $\hat{G}$ and exponents $\varepsilon_i = \pm 1$, $i = 1,\ldots, m$. Then H is a subgroup of K and we shall show that H actually is the whole of K. The proof will proceed by induction on ν the total number of vertices lying in R. If $\nu = 1$, then R consists of a single point and this case was considered above.

Suppose now that R contains at least one polygon. Then by reordering the sequence $P_1,\ldots, P_k$, it may be assumed that R contains a polygon Q_1 adjacent to P_k as in Figure 12.

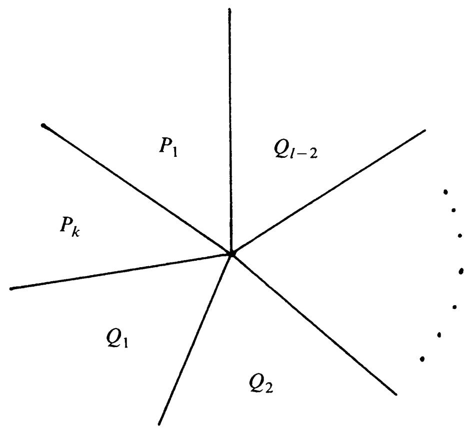


FIG. 12.

The sequence $Q_1, Q_2, \ldots, Q_{l-1} = P_1, Q_l = P_k$ corresponds to the element $q = \hat{y}_1\hat{y}_2 \cdots \hat{y}_l$ say of K and hence of H by the above. Because the sequence $P_1, P_2, \ldots, P_k, Q_1, Q_2, \ldots, Q_{l-2}$ contains fewer polygons in its inside, it corresponds by induction to an element r of H. We can calculate $\hat{x}_1 \cdots \hat{x}_k$ in terms of q and r as follows:

The sequence $P_1, P_2, \ldots, P_k, Q_1, Q_2, \ldots, Q_{l-2}$ gives rise to the word r as follows:

$$\begin{aligned}
P_k &= \hat{x}_2 \cdots \hat{x}_k P_1 \\
Q_1 &= \hat{y}_1 P_k = \hat{y}_1 \hat{x}_2 \cdots \hat{x}_k P_1 \\
Q_2 &= \hat{y}_l Q_1 = \hat{y}_l \hat{y}_1 \hat{x}_2 \cdots \hat{x}_k P_1 \\
Q_3 &= \hat{y}_{l-1} Q_2 = \hat{y}_{l-1} \hat{y}_l \hat{y}_1 \hat{x}_2 \cdots \hat{x}_k P_1 \\
&\ \vdots \\
Q_{l-2} &= \hat{y}_4 \cdots \hat{y}_l \hat{y}_1 \hat{x}_1 \cdots \hat{x}_k P_1 \\
P_1 &= \hat{y}_3 \cdots \hat{y}_l \hat{y}_1 \hat{x}_2 \cdots \hat{x}_k P_1
\end{aligned}$$

So
$$r = \hat{y}_2^{-1} \hat{y}_1^{-1} q \hat{y}_1 \hat{x}_1^{-1} \hat{x}_1 \hat{x}_2 \cdots \hat{x}_k.$$

Using the fact that $\hat{y}_2 = \hat{x}_1^{-1}$ we see that

$$\hat{x}_1 \cdots \hat{x}_k = (y_1 y_2)^{-1} q^{-i} (y_1 y_2) r.$$

By induction q and r lie in H. Since H is a normal subgroup of $\hat{G}$ it follows that $\hat{x}_1\hat{x}_2 \cdots \hat{x}_k$ also lies in H.

If R contains no polygons, then a similar argument works by pushing in R along edges.

In group theoretic language G is generated by $a_1, b_1, a_2, b_2, \ldots, a_\gamma, b_\gamma$ and has one relation

$$c = [a_1, b_1][a_2, b_2] \cdots [a_\gamma, b_\gamma] = 1.$$

7. The Orbit Space S_G. By analogy with the torus considered in the introduction, let S_G be the hyperbolic plane divided out by the action of the group G. Then S_G is obtained from the polygon P by identifying edges $e_1 = Z_i Z_{i+1}$ with $e_2 = Z_j Z_{j+1}$ if there is a group element transporting e_1 into e_2. Now to any edge e there is a unique disjoint edge f and a unique generator or its inverse in $X \cup X^{-1}$ which maps e to f. Conversely associated with each generator, there is a unique pair of edges which are identified by this generator. The result of identifying P by these generators is the orientable surface of genus γ. So it only remains to show that these are the only identifications.

Let e and f be two edges of P and g an element of G which takes e into f. Let xP be the

polygon having the edge f in common with P where $x \in X \cup X^{-1}$. Then gP is either P or xP. But we have seen that there is a 1-1 correspondence between the polygons gP and the group elements g. Hence either $g = 1$ or $g = x$. So the only identifications are the ones above. These identifications define the orientable surface of genus γ as in Massey's book [5].

8. All Elements of $G - \{1\}$ Are Hyperbolic. Suppose that g is a nontrivial element of G which is not hyperbolic. Then there would be points Z, gZ in $\mathcal{H}$ arbitrarily close to one another. However, there is a positive minimum for the distances between any two points in P which are to be identified, and this is therefore a minimum for all identified pairs in $\mathcal{H}$. Therefore all elements of $G - \{1\}$ are hyperbolic.

9. Acknowledgements. I would like to acknowledge some particularly helpful conversations with Peter Scott, especially one during a wet, blustery walk over the Sussex Downs. I would also like to thank David Singerman for a helpful comment and Joe Taylor whose magic computer helped with some of the diagrams.

References

1. C. Caratheodory, Theory of Functions of a Complex Variable, Chelsea, 1950.
2. H. S. M. Coxeter, Non Euclidean Geometry, University of Toronto Press, 1942.
3. H. Guggenheimer, Plane Geometry and Its Groups, Holden-Day 1967.
4. W. Magnus, Non Euclidean Tesselations and Their Groups, Academic Press, 1974.
5. W. Massey, Algebraic Topology, Harbrace, 1968.
6. H. Poincaré, Les géométries non-euclidiennes, Revue Génerale des Sciences, no. 23, 2 (1981).
7. B. Riemann, Uber die Hypothesen, welche der Geometrie zu Grunde Liegen, Abhandlungen der Koniglichen Gesellschaft der Wissenschaften zu Gottingen, 13 (1866) 254–268.
8. E. Scholz, Geschichte des Mannig faltigkeitsbeariffs von Riemann bis Poincaré, Birkhausen, 1980.
9. C. L. Siegel, Topics in Complex Function Theory, vol. 2, Wiley Interscience, 1977.
10. W. Thurston, Lecture Notes on the Geometry and Topology of 3-Manifolds, Princeton University.

A NECESSARY AND SUFFICIENT CONDITION FOR THE PRIMALITY OF FERMAT NUMBERS

RICHARD MCINTOSH
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada, T2N 1N4

A Fermat number is an integer of the form

$$F(n) = 2^{2^n} + 1, \qquad n \geqslant 0.$$

If $F(n)$ is prime, it is called a Fermat prime. The purpose of this paper is to give a necessary and sufficient condition for Fermat primes.

We will need the tangent numbers $T(n)$ and the Bernoulli numbers $B(n)$. They are defined to be the coefficients in the following power series

$$\tan z = \sum_{n=0}^{\infty} T(n) z^n/n!, \quad z/(e^z - 1) = \sum_{n=0}^{\infty} B(n) z^n/n!.$$

THEOREM. *$F(n)$ is prime if and only if $F(n)$ does not divide $T(F(n) - 2)$.*

Richard McIntosh: I am an undergraduate student. I discovered the theorem when I was working on a summer research project at the University of Calgary in 1980.

polygon having the edge f in common with P where $x \in X \cup X^{-1}$. Then gP is either P or xP. But we have seen that there is a 1-1 correspondence between the polygons gP and the group elements g. Hence either $g = 1$ or $g = x$. So the only identifications are the ones above. These identifications define the orientable surface of genus γ as in Massey's book [5].

8. All Elements of $G - \{1\}$ Are Hyperbolic. Suppose that g is a nontrivial element of G which is not hyperbolic. Then there would be points Z, gZ in $\mathcal{H}$ arbitrarily close to one another. However, there is a positive minimum for the distances between any two points in P which are to be identified, and this is therefore a minimum for all identified pairs in $\mathcal{H}$. Therefore all elements of $G - \{1\}$ are hyperbolic.

9. Acknowledgements. I would like to acknowledge some particularly helpful conversations with Peter Scott, especially one during a wet, blustery walk over the Sussex Downs. I would also like to thank David Singerman for a helpful comment and Joe Taylor whose magic computer helped with some of the diagrams.

References

1. C. Caratheodory, Theory of Functions of a Complex Variable, Chelsea, 1950.
2. H. S. M. Coxeter, Non Euclidean Geometry, University of Toronto Press, 1942.
3. H. Guggenheimer, Plane Geometry and Its Groups, Holden-Day 1967.
4. W. Magnus, Non Euclidean Tesselations and Their Groups, Academic Press, 1974.
5. W. Massey, Algebraic Topology, Harbrace, 1968.
6. H. Poincaré, Les géométries non-euclidiennes, Revue Génerale des Sciences, no. 23, 2 (1981).
7. B. Riemann, Uber die Hypothesen, welche der Geometrie zu Grunde Liegen, Abhandlungen der Koniglichen Gesellschaft der Wissenschaften zu Gottingen, 13 (1866) 254–268.
8. E. Scholz, Geschichte des Mannig faltigkeitsbeariffs von Riemann bis Poincaré, Birkhausen, 1980.
9. C. L. Siegel, Topics in Complex Function Theory, vol. 2, Wiley Interscience, 1977.
10. W. Thurston, Lecture Notes on the Geometry and Topology of 3-Manifolds, Princeton University.

A NECESSARY AND SUFFICIENT CONDITION FOR THE PRIMALITY OF FERMAT NUMBERS

RICHARD MCINTOSH
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada, T2N 1N4

A Fermat number is an integer of the form

$$F(n) = 2^{2^n} + 1, \qquad n \geqslant 0.$$

If $F(n)$ is prime, it is called a Fermat prime. The purpose of this paper is to give a necessary and sufficient condition for Fermat primes.

We will need the tangent numbers $T(n)$ and the Bernoulli numbers $B(n)$. They are defined to be the coefficients in the following power series

$$\tan z = \sum_{n=0}^{\infty} T(n) z^n/n!, \quad z/(e^z - 1) = \sum_{n=0}^{\infty} B(n) z^n/n!.$$

THEOREM. *$F(n)$ is prime if and only if $F(n)$ does not divide $T(F(n) - 2)$.*

Richard McIntosh: I am an undergraduate student. I discovered the theorem when I was working on a summer research project at the University of Calgary in 1980.

Proof. The definition of $\tan z$ implies

$$\tan z = \frac{(e^{iz} - e^{-iz})}{i(e^{iz} + e^{-iz})} = \frac{1}{z}\left(\frac{2iz}{e^{2iz}} - iz\right)$$
$$= \frac{1}{z}\left(\frac{2iz}{e^{2iz} - 1} - \frac{4iz}{e^{4iz} - 1} - iz\right)$$
$$= \frac{1}{z}\left(-iz + \sum_{n=0}^{\infty}\left[(2iz)^n - (4iz)^n\right]B(n)/n!\right);$$

and by equating coefficients we obtain the identity

$$(*) \qquad T(2n-1) = 2^{2n-1}(2^{2n}-1)|B(2n)|/n, \qquad n \geqslant 1.$$

Let $N(2n)$ and $D(2n)$ denote the numerator and denominator of $|B(2n)|$, respectively. $(N(2n), D(2n)) = 1$. From the von Staudt-Clausen theorem [2]

$$D(2n) = \prod_{\substack{p \text{ prime};\\ (p-1)|2n}} p.$$

It follows from (*) that

$$T(F(n)-2) = \frac{2^{F(n)-2}(2^{F(n)-1}-1)N(F(n)-1)}{2^{2^n-1}D(F(n)-1)},$$
$$D(F(n)-1) = \prod_{\substack{p \text{ prime};\\ (p-1)|2^{2^n}}} p = 2\tilde{F}(n)\tilde{F}(n-1)\cdots\tilde{F}(0),$$

where $\tilde{F}(k) = \begin{cases} F(k) & \text{if } F(k) \text{ is prime,}\\ 1 & \text{if } F(k) \text{ is composite.}\end{cases}$

Thus

$$T(F(n)-2) = \frac{F(2^n-1)F(2^n-2)\cdots F(0)2^{F(n)-2^n-2}N(F(n)-1)}{\tilde{F}(n)\tilde{F}(n-1)\cdots\tilde{F}(0)}.$$

Since $(D(F(n)-1), N(F(n)-1)) = 1$ and $\tilde{F}(n)$ is odd, it follows that

$$\left(\tilde{F}(n), 2^{F(n)-2^n-2}N(F(n)-1)\right) = 1.$$

$F(k+1) = 2 + F(k)F(k-1)\cdots F(0)$ for $k \geqslant 0$ implies $(F(n), F(k)) = 1$ for all $k \neq n$. Hence $F(n)$ is prime if and only if $F(n)$ does not divide $T(F(n)-2)$, which completes the proof.

References

1. Thomas Buckholtz and Donald Knuth, Computation of tangent, Euler, and Bernoulli numbers, Math. Comp., 21 (1967) 663–688.

2. Hans Rademacher, Topics in Analytic Number Theory, Springer-Verlag, Berlin, 1973.

MISCELLANEA

92. The greatest enemy to true arithmetic work is found in so-called practical or illustrative problems, which are freely given to our pupils, of a degree of difficulty and complexity altogether unsuited to their age and mental development...

—F. A. Walker, 1899 (quoted in R. E. Moritz, *Miscellanea Mathematica*, New York, 1914, p. 81).

Proof. The definition of tan z implies

$$\tan z = \frac{(e^{iz} - e^{-iz})}{i(e^{iz} + e^{-iz})} = \frac{1}{z}\left(\frac{2iz}{e^{2iz}} - iz\right)$$
$$= \frac{1}{z}\left(\frac{2iz}{e^{2iz} - 1} - \frac{4iz}{e^{4iz} - 1} - iz\right)$$
$$= \frac{1}{z}\left(-iz + \sum_{n=0}^{\infty}\left[(2iz)^n - (4iz)^n\right]B(n)/n!\right);$$

and by equating coefficients we obtain the identity

$$(*) \qquad T(2n-1) = 2^{2n-1}(2^{2n}-1)|B(2n)|/n, \qquad n \geqslant 1.$$

Let $N(2n)$ and $D(2n)$ denote the numerator and denominator of $|B(2n)|$, respectively. $(N(2n), D(2n)) = 1$. From the von Staudt-Clausen theorem [**2**]

$$D(2n) = \prod_{\substack{p \text{ prime;} \\ (p-1)|2n}} p.$$

It follows from (*) that

$$T(F(n)-2) = \frac{2^{F(n)-2}(2^{F(n)-1}-1)N(F(n)-1)}{2^{2^n-1}D(F(n)-1)},$$
$$D(F(n)-1) = \prod_{\substack{p \text{ prime;} \\ (p-1)|2^{2^n}}} p = 2\tilde{F}(n)\tilde{F}(n-1)\cdots\tilde{F}(0),$$

where $\tilde{F}(k) = \begin{cases} F(k) & \text{if } F(k) \text{ is prime,} \\ 1 & \text{if } F(k) \text{ is composite.}\end{cases}$

Thus

$$T(F(n)-2) = \frac{F(2^n-1)F(2^n-2)\cdots F(0)2^{F(n)-2^n-2}N(F(n)-1)}{\tilde{F}(n)\tilde{F}(n-1)\cdots\tilde{F}(0)}.$$

Since $(D(F(n)-1), N(F(n)-1)) = 1$ and $\tilde{F}(n)$ is odd, it follows that

$$\left(\tilde{F}(n), 2^{F(n)-2^n-2}N(F(n)-1)\right) = 1.$$

$F(k+1) = 2 + F(k)F(k-1)\cdots F(0)$ for $k \geqslant 0$ implies $(F(n), F(k)) = 1$ for all $k \neq n$. Hence $F(n)$ is prime if and only if $F(n)$ does not divide $T(F(n)-2)$, which completes the proof.

References

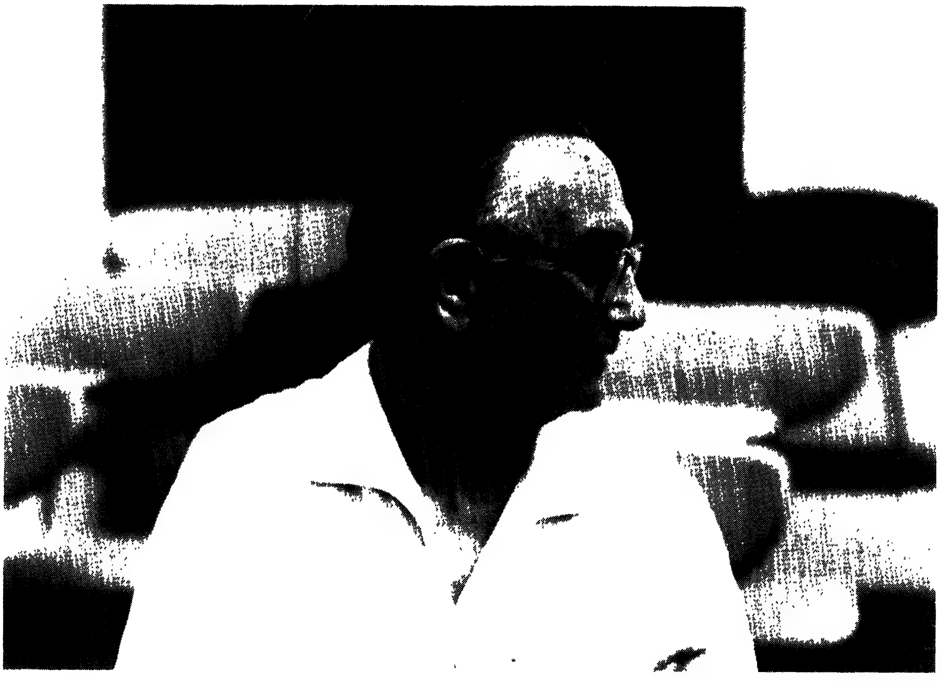
1. Thomas Buckholtz and Donald Knuth, Computation of tangent, Euler, and Bernoulli numbers, Math. Comp., 21 (1967) 663–688.

2. Hans Rademacher, Topics in Analytic Number Theory, Springer-Verlag, Berlin, 1973.

MISCELLANEA

92. The greatest enemy to true arithmetic work is found in so-called practical or illustrative problems, which are freely given to our pupils, of a degree of difficulty and complexity altogether unsuited to their age and mental development...

—F. A. Walker, 1899 (quoted in R. E. Moritz, *Miscellanea Mathematica*, New York, 1914, p. 81).



What ever happened to proofs by transfinite induction? See p. 145.

showing cute examples of how to mislead others with the misuse of statistics. The CUPM panel's recommendation that this course be taught by an experienced senior faculty member is very important. Furthermore, I recommend that this person have experience in teaching statistics.

Taking a broader view, one sees that those at the forefront of statistics are beginning to discuss "Statistics as a Discipline." In fact, this was the title of the main address presented by Paul Minton at the second annual Conference for Texas Statisticians held at the Baylor University (April, 1982). The main point of Professor Minton's talk concerned the idea of offering *undergraduate degrees* in statistics.

There are two points I want to make. First, it is conceivable that a statistics course "for coping with life" could cover most of the topics in the list given by the CUPM panel. Second, topic 14 in a list of topics for this statistics course could be "mathematics and its dangers." Need I say more?

Since it would be cowardly to avoid a recommendation for a replacement for topic 14, I suggest:

14. The Use and Interpretation of Statistics in Everyday Life.

Danny W. Turner
Department of Mathematics
Baylor University
Waco, Texas 76798

MISCELLANEA

With Apologies to John Milton

94. John Milton wrote a famous sonnet on being blind. This modern version looks at a different kind of blindness.

When I consider how my grants are spent,
Ere half my days in these dark halls and wide;
My teaching talent, which I'd love to hide,
Lodged with me useless, though my soul more bent
To serve therewith my Provost, and present
My teaching ratings, lest he returning chide;
"Do Deans exact much teaching, grants denied?"
I fondly ask: the Provost, to prevent
That murmur, soon replies, "Deans do not need
Either that teaching or those grants. Who best
Does his committee work, serves best. His state
Is Deanly: thousands at his bidding speed
And fly to Washington, disdaining rest;
They also serve who only loaf and wait."

—Edwin Hewitt

ANSWER TO "PHOTO" ON PAGE 100

The subject is Max Zorn; the picture was taken in the summer of 1981. He is alive and well and very much a part of the mathematical community in Bloomington, Indiana.

RELATIONAL DATABASES

QUENTIN F. STOUT
Mathematical Sciences, State University of New York, Binghamton, NY 13901

PATRICIA A. WOODWORTH
Department of Mathematics, Ithaca College, Ithaca, NY 14850

1. Introduction. A cursory view of employment advertisements shows that database management is an area of intense current interest. Informally, a database is just a collection of programs and data sets which store and retrieve large amounts of information in an orderly manner. Databases are widely used in industry and government, and are increasingly useful for research assistance. For example, a popular system called DIALOG, operated by Lockheed Missile and Space Company, Inc., offers researchers telephone access, via computer terminals, to over 100 separate databases containing abstracts, key words, cross-indices, etc. One of these is SCISEARCH, constructed by the Institute for Scientific Information, which essentially contains the *Science Citation Index* in a computer-accessible form. SCISEARCH contains about 3 million entries, and can easily answer queries such as "List all papers written in English which reference either B. Sz.-Nagy's paper *Sur les contractions de l'espace de Hilbert*, Acta. Sci. Math. (Szeged), 15 (1963) 82–85, or J. J. Schaffer's paper *On unitary dilations of contractions*, Proc. Amer. Math. Soc., 6 (1955) 322." The actual request sent to SCISEARCH could be

```
SELECT     CR = NAGY BS, 1963, V15, P82
SELECT     CR = SCHAFFER JJ, 1955, V6, P322
COMBINE    1 or 2
LIMIT      3/ENG
```

(This form is not the most desirable, but is easily learned.) DIALOG will also contain the American Mathematical Society's MATHFILE database, built from the reviews appearing in *Mathematical Reviews* [**30**].

Besides being useful research tools, databases are an interesting source of mathematical problems. These include statistical and cryptographic considerations of security [**28**, Chap. 9], questions of computational complexity [**4**], [**28**, Chap. 6], and logical and algebraic questions of how to store and access the information. It is this last area which is the focus of this paper. There are several competing schemes for database organization, one of which is called the relational form. This is the form which has been most intensively studied, and was advanced by Codd [**9**] as being sufficiently powerful to model real-world situations while being conceptually and computationally simple enough to manipulate. It is particularly susceptible to mathematical analysis, which may account for its popularity in the research community. We feel that an expository account of some of the issues leading to, and arising from, the so-called normal forms for relational databases may be of interest to mathematicians, all of whom have encountered relations

---

Quentin F. Stout: My 1977 Ph.D. was in operator theory and functional analysis, directed by John B. Conway at Indiana University. Since 1976 I have been an Assistant Professor at State University of New York at Binghamton. I have long had an interest in both computer science and mathematics, and I especially enjoy problems that draw from both disciplines. My current research is primarily in computer science, particularly massively parallel computation.

Patricia A. Woodworth: I received my Ph.D. from the State University of New York at Binghamton in 1977 under David Klarner. My dissertation dealt with the combinatorial problem of plane tree enumeration. Since then, I have been an Assistant Professor in the Mathematics Department at Wilkes College (one year) and Ithaca College. Recently I have become increasingly involved in computer science education and research.

C#	CN	SEC	CR	PN	P#
111	Calculus I	1	4	Jones	125
111	Calculus I	2	4	Smith	382
111	Calculus I	3	4	Thomas	418
111	Calculus I	4	4	Jones	125
111	Calculus I	5	4	Johnson	432
112	Calculus II	1	4	Smith	153
112	Calculus II	2	4	Thomas	418
243	Linear Algebra	1	3	Jones	125
243	Linear Algebra	2	3	Smith	382
243	Linear Algebra	3	3	Johnson	432
318	Analysis	1	3	Smith	153

TOTAL-OFFERINGS

FIG. 1

C#	CN	SEC	CR	PN	P#
111	Calculus I	3	4	Thomas	418
112	Calculus II	2	4	Thomas	418

TOTAL-OFFERINGS_p

CN
Calculus I
Calculus II

$\text{TOTAL-OFFERINGS}_p\ [\text{CN}]$

FIG. 2

in several courses but few of whom have used them in nontrivial applications. This paper concerns only a small, though significant, portion of database design. Readers who wish to obtain a broader view are encouraged to consult such texts as Cardano [**8**], Date [**14**], Martin [**24**] or Ullman [**28**].

As an example, suppose DB University wishes to computerize its information concerning courses. The information they wish to keep includes course name (CN), course number (C#), section (SEC), credit (CR), professor name (PN), and professor number (P#). CN, C#, SEC, CR, PN, P# are called *attributes*, which are just sets of all possible values. For example, P# may denote the set of all 3 digit numbers. At any instant the data is a relation on {CN, C#, SEC, CR, PN, P#}, i.e., a subset of CN $\times$ C# $\times$ SEC $\times$ CR $\times$ PN $\times$ P#. The role of time is quite important, since the relation can change from instant to instant. However, there are some features, beyond the attribute set itself, that remain constant. For example, we will always assume that P# uniquely determines PN, that C# uniquely determines CN and CR, that CN uniquely determines C#, and that C# and SEC together uniquely determine P#. A possible instance of this relation, which we will call TOTAL-OFFERINGS, is given in Fig. 1. The tabular form of Fig. 1 is particularly useful for answering queries such as "What are the names of the courses taught by professor number 418?" Such queries can be answered using the relational operators *selection* and *projection*. We use the shorthand notation "P# = 418" to denote the predicate "The P#th component of X equals 418" for $X \in$ CN $\times$ C# $\times$ SEC $\times$ CR $\times$ PN $\times$ P#. If we let P denote this predicate, then TOTAL-OFFERINGS$_{\mathrm{P}}$ denotes the subrelation consisting of all elements X of TOTAL-OFFERINGS such that P(X) is true. We then project TOTAL-OFFERINGS$_{\mathrm{P}}$ onto the attribute CN, denoted by TOTAL-OFFERINGS$_{\mathrm{P}}$[CN], to obtain our answer. This projection is just the standard projection of a subset of a crossproduct onto one of its components. See Fig. 2.

Despite its utility, TOTAL-OFFERINGS has some undesirable defects. For example, five different elements of the relation contain the information that the course numbered 111 is named CALCULUS I and carries 4 credits. There are several problems here. In addition to being an inefficient use of space, the relation is more susceptible to consistency errors. That is, if one inserts a new section of Calculus I and enters 3 as the number of credits, then the new entry will be inconsistent with the previous ones. Cross checking can be done to prevent this but it is time-consuming and usually not done. We can solve both problems by decomposing the TOTAL-OFFERINGS relation. We introduce a relation COURSES, where COURSES = TOTAL-OFFERINGS [C#, CN, CR], and a temporary relation $R1$, where $R1$ = TOTAL OFFERINGS [C#, SEC, PN, P#]. These are given in Fig. 3. By storing COURSES and $R1$ instead of TOTAL-OFFERINGS, we have reduced the total space needed to store the information, eliminated some redundancy, and reduced the possibility of certain logical errors. Now to introduce a sixth section of CALCULUS I we only need to modify $R1$. We can proceed further noting that we have a similar situation with respect to the professors. We decompose $R1$ into PROFESSORS = $R1$[P#, PN] and OFFERINGS = $R1$[C#, SEC, P#], illustrated in Fig. 4. The properties of projections insure that PROFESSORS = TOTAL-OFFERINGS [P#, PN] and OFFERINGS = TOTAL-OFFERINGS [C#, SEC, P#].

Each of COURSES, PROFESSORS, and OFFERINGS is a clearly defined logical unit to which we have assigned a natural name, and so they provide a reasonable way to compress the information contained in TOTAL-OFFERINGS. The collection {COURSES, PROFESSORS, OFFERINGS} is called a *decomposition* of TOTAL-OFFERINGS, with decomposition being an informal term meaning that the original relation has been replaced by a collection of smaller ones. Further, we claim that the three subrelations in the decomposition contain all the information of the original. To define precisely what this means we add the notion of *join*, the third fundamental operation on relations. If $R1$ and $R2$ are relations with attribute sets $\mathcal{Q}_1$ and $\mathcal{Q}_2$, then the join of $R1$ and $R2$, denoted $R1 * R2$, has attribute set $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2$ and is given by

$$R1 * R2 = \{t \in \otimes\mathcal{Q} : \exists t_1 \in R1, t_2 \in R2 \ni t[\mathcal{Q}_1] = t_1 \quad \text{and} \quad t[\mathcal{Q}_2] = t_2\}$$

C#	CN	CR
111	Calculus I	4
112	Calculus II	4
243	Linear Algebra	3
318	Analysis	3

COURSES = TOTAL-OFFERINGS [C#,CN,CR]

C#	SEC	PN	P#
111	1	Jones	125
111	2	Smith	382
111	3	Thomas	418
111	4	Jones	125
111	5	Johnson	432
112	1	Smith	153
112	2	Thomas	418
243	1	Jones	125
243	2	Smith	382
243	3	Johnson	432
318	1	Smith	153

R1 = TOTAL-OFFERINGS [C#,SEC,PN,P#]

FIG. 3

where we use $t[\mathcal{A}_i]$ to mean the projection of t onto $\mathcal{A}_i$. It is easy to show that the join is commutative, associative, and, if $\mathcal{A}_1 \cap \mathcal{A}_2 = \phi$, is just the cross product. Our claim that the decomposition contains all of the information of TOTAL-OFFERINGS is the claim that OFFERINGS∗COURSES∗PROFESSORS = TOTAL-OFFERINGS. {OFFERINGS, COURSES, PROFESSORS} is said to be a *lossless decomposition* of TOTAL-OFFERINGS, where in general if R is a relation with attribute set $\mathcal{A}$ and $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n$, then $\{R[\mathcal{A}_1],\ldots, R[\mathcal{A}_n]\}$ is a lossless decomposition of R if and only if $R = R[\mathcal{A}_1] * \cdots * R[\mathcal{A}_n]$. A decomposition which is not lossless is said to be *lossy*. What is lost in a lossy decomposition is the information that certain combinations are impossible, since we always have $R[\mathcal{A}_1] * \cdots * R[\mathcal{A}_n] \supseteq R$. For example, TOTAL-OFFERINGS [C#, PN]∗TOTAL-OFFERINGS [PN, P#] has such entries as (112, SMITH, 382). With OFFERINGS∗COURSES∗PROFESSORS this cannot happen, for we assume that P# uniquely determines PN and C# uniquely determines CN. (The next section

P#	PN
125	Jones
382	Smith
418	Thomas
432	Johnson
153	Smith

PROFESSORS = Rl [P#,PN]

C#	SEC	P#
111	1	125
111	2	382
111	3	418
111	4	125
111	5	432
112	1	153
112	2	418
243	1	125
243	2	382
243	3	432
318	1	153

OFFERINGS = Rl [C#,SEC,P#]

FIG. 4

contains a proof that this decomposition is lossless.) OFFERINGS contains all of the significant information about the correspondence between courses, sections, and professors, with the other two relations expanding some details.

Lossless decompositions can give space efficiency and reduce insertion anomalies, as we have seen, and can reduce certain other anomalies. For example, suppose that next semester analysis is not being taught. Then that semester TOTAL-OFFERINGS would contain no information about analysis. In the decomposed relations, OFFERINGS will also not include any information concerning analysis, but in COURSES we can keep the information that analysis is course number 318. It is still true that the join of the subrelations will be the entire relation, but now the subrelations can contain additional information, and COURSES $\supseteq$ TOTAL-

OFFERINGS [C#, CN, CR]. The decomposed version allows us to avoid a deletion anomaly whereby a deletion of some information (analysis is not being taught) requires us to delete other information (analysis is course 318). However, this possibility of having the whole be less than the sum of its parts causes some dissension, with several researchers insisting (perhaps only implicitly) that at all times each subrelation is a projection of some, perhaps never actually constructed, universal relation. This is usually called the *universal relation assumption* **[5]**, an assumption we shall not make.

Lossless decompositions can also help avoid update anomalies, in which a single change forces other changes. For example, if professor number 125 changes his or her name to ROBERTS, then that would necessitate changing 3 entries in TOTAL-OFFERINGS. Changing only one entry would introduce an inconsistency. In the decomposed version this would not happen since only one entry of PROFESSORS would be affected.

The universal relation is quite powerful, but the problem of insertion, deletion, or update anomalies makes it unsuitable, pushing us to decompose into more fundamental units from which we can reconstruct the universal relation by the join operator. Codd **[11]** stated that the two most important reasons for using decompositions are:

(1) To reduce the need for restructuring the collection of relations as new types of data are introduced;
(2) To reduce the incidence of undesirable insertion, update and deletion anomalies.

We shall see that the normal forms essentially accomplish this, although other considerations will introduce conflicting goals.

2. Second and Third Normal Form. The normal forms, introduced by Codd **[10]**, formalize the conditions a relation must satisfy in order to avoid the redundancies and anomalies described in the introduction. In this section, we will define the first three normal forms and the process of decomposing a relation into normal forms. In the relation TOTAL-OFFERINGS (see Fig. 1) a redundancy resulted because the value of the attribute CR depended on the value of the attribute C#. For example, if the value of C# is 112, then, at this point in time, the value of CR is 4. For a given value of C# the value of CR may change over time but, at any instant, if the value of C# is known, then there is only one possible value of CR.

The dependency of CR on C# is called a functional dependency. In general, assume R is a relation with attribute set $\mathcal{Q}$ and assume X and Y are subsets of $\mathcal{Q}$. In R, Y is *functionally dependent* on X (*X functionally determines Y*), written as $X \to Y$, if $r_1, r_2 \in R$ and $r_1[X] = r_2[X]$ implies that $r_1[Y] = r_2[Y]$. $X \leftrightarrow Y$ means $X \to Y$ and $Y \to X$. As specified in the introduction, the attributes of the relation TOTAL-OFFERINGS satisfy the following functional dependencies:

$$\begin{array}{ll} \mathrm{C\#} \leftrightarrow \mathrm{CN} & \mathrm{C\#}, \mathrm{SEC} \to \mathrm{P\#} \\ \mathrm{C\#} \to \mathrm{CR} & \mathrm{P\#} \to \mathrm{PN} \end{array}$$

It is important to realize that a functional dependency cannot be determined from any particular instance of a relation. For example, if a particular instance of TOTAL-OFFERINGS contained only one section of each course, then it would appear that C# $\to$ SEC. Since this is not true for all instances of TOTAL-OFFERINGS, this is an invalid conclusion. Therefore the functional dependencies must be specified by a person familiar with the situation being modeled. An added complication is that these dependencies can also be changed over time. For example, a curriculum committee may decide that the credit for a course may be determined by the course number and the level of the student taking the course. This complication will be forbidden here.

It can easily be shown that functional dependencies satisfy the following properties:

FD1. Projectivity: If $Y \subseteq X$, then $X \to Y$.
e.g., $\{\mathrm{C\#}, \mathrm{SEC}\} \to \mathrm{C\#}$
FD2. Additivity: If $X \to Y$ and $X \to Z$, then $X \to Y \cup Z$.
e.g., $\mathrm{C\#} \to \mathrm{CN}$ and $\mathrm{C\#} \to \mathrm{CR} \Rightarrow \mathrm{C\#} \to \{\mathrm{CN}, \mathrm{CR}\}$

FD3. Transitivity: If $X \to Y$ and $Y \to Z$, then $X \to Z$.
e.g., {C#, SEC} $\to$ P# and P# $\to$ PN $\Rightarrow$ {C#, SEC} $\to$ PN

FD4. Augmentation: If $X \to Y$, then $X \cup Z \to Y$.
e.g., C# $\to$ CR $\Rightarrow$ {C#, SEC} $\to$ CR

FD5. Decomposition: If $X \to Y$ and $Z \subseteq Y$, then $X \to Z$.
e.g., C# $\to$ {CN, CR} $\Rightarrow$ C# $\to$ CN

Rule FD3 corresponds to the composition of functions. FD1 implies that $X \to X$ for any set X of attributes.

A *key* for a relation R is a subset X of the attributes $\mathcal{Q}$ of R that satisfies the following two properties:

K1: $X \to A$ for all $A \in \mathcal{Q}$.

K2: No proper subset of X has property K1.

There exists at least one key for any relation since the set of all attributes vacuously satisfies K1, and hence at least one of its subsets satisfies both K1 and K2. By K1, the values of the attributes in a key uniquely determine the values of all other attributes in the relation. The attributes {C#, SEC} are a key for the relation TOTAL-OFFERINGS since CN, CR, P#, PN are all functionally dependent on C# while PN and P# are not functionally dependent on either C# or SEC separately. Since CN $\leftrightarrow$ C#, it is easy to see that {CN, SEC} is also a key for TOTAL-OFFERINGS.

An attribute is *prime* if it is a member of some key. Otherwise it is *nonprime*. In TOTAL-OFFERINGS, C#, CN, SEC are prime while CR, P#, PN are nonprime.

It is now possible to define Codd's normal forms. (As defined here, a relation is automatically in first normal form. A relation is in first normal from (1NF) if none of the attributes have elements which are themselves sets.)

Assume R is a relation, X is a subset of the attributes of R and A is a *nonprime* attribute which is *not* in X.

2NF: R is in *second normal form* if $X \to A$ implies that X is not properly contained in any key of R.

3NF: R is in *third normal form* if $X \to A$ implies that X contains some key of R.

If X contains a key, then it is not possible for X to be contained in a key (condition K2), and therefore 3NF implies 2NF. TOTAL-OFFERINGS is neither a 3NF nor a 2NF relation since C# $\to$ CR and C# is contained in the key {C#, SEC}.

The goal is to decompose a relation R into a collection of 3NF relations $R_1, \ldots, R_n$ which satisfies the lossless decomposition property (i.e., $R_1 * \cdots * R_n = R$). The desirability of a lossless decomposition was indicated in the introduction. The following theorem can be used to achieve this goal.

THEOREM 1. *Assume R is a relation with attributes $\mathcal{Q}$. Let $X, Y \subset \mathcal{Q}$ and let $Z = \mathcal{Q} - (X \cup Y)$. If $X \to Y$, then the 2 projections $R_1 = R[X \cup Y]$ and $R_2 = R[X \cup Z]$ form a lossless decomposition of R.*

Proof.

$$R_1 * R_2 = \{(x, y, z) : x \in X, y \in Y, z \in Z, (x, y) \in R_1 \text{ and } (x, z) \in R_2\}.$$

Clearly $R \subseteq R_1 * R_2$. Assume $(x, y, z) \in R_1 * R_2$. Since $(x, z) \in R_2 = R[X \cup Z]$, there exists a $y' \in Y$ such that $(x, y', z) \in R$. This implies $(x, y') \in R_1 = R[X \cup Y]$. But $(x, y) \in R_1$, $(x, y') \in R_1$, $X \to Y$ implies that $y = y'$. Therefore $(x, y, z) = (x, y', z) \in R$. So $R_1 * R_2 \subseteq R$. □

Note that this theorem provides a technique to generate a lossless decomposition but does not immediately guarantee a decomposition into 3NF relations.

To see how this theorem can be used to generate a 3NF decomposition, consider again the relation TOTAL-OFFERINGS. This relation is not in 3NF. One problem, as determined earlier,

C#	CR
111	4
112	4
243	3
318	3

O_1 = TOTAL-OFFERINGS [C#,CR]

P#	PN
125	Jones
382	Smith
418	Thomas
432	Johnson
153	Smith

O_3 = TOTAL-OFFERINGS [C#,CN,SEC,P#]

C#	CN	SEC	P#
111	Calculus I	1	125
111	Calculus I	2	382
111	Calculus I	3	418
111	Calculus I	4	125
111	Calculus I	5	432
112	Calculus II	1	153
112	Calculus II	2	418
243	Linear Algebra	1	125
243	Linear Algebra	2	382
243	Linear Algebra	3	432
318	Analysis	1	153

O_4 = TOTAL-OFFERINGS [C#,CN,SEC,P#]

FIG. 5

is due to the "bad" dependency C# $\to$ CR where C# does not contain a key (CR is a nonprime attribute). Theorem 1 guarantees that the decomposition of TOTAL-OFFERINGS into the two projections 0_1 = TOTAL-OFFERINGS [C#, CR] and 0_2 = TOTAL-OFFERINGS [C#, CN, SEC, P#, PN] is lossless. In fact, 0_1 is now a 3NF relation. But 0_2 is not since P# $\to$ PN and P# does not contain a key. ({C#, SEC} and {CN, SEC} are still keys for 0_2.) Then applying Theorem 1 to the relation 0_2 using the "bad" dependency P# $\to$ PN yields the lossless decomposition of 0_2 into $0_3 = 0_2$[P#, PN] and $0_4 = 0_2$[C#, CN, SEC, P#] (see Fig. 5). This time, both relations 0_3 and 0_4 are 3NF since no nonprime elements are dependent on any set that does not contain a key. The three relations $0_1, 0_3, 0_4$ form a lossless 3NF decomposition of TOTAL-OFFERINGS.

The following theorem guarantees that this technique will always yield a desired decomposition for any relation.

THEOREM 2. *If a relation R is not in* 3NF, *then it is possible to decompose it losslessly into a collection $R_1, \ldots, R_n$ of* 3NF *relations.*

Proof. Assume R is a relation with attributes $\mathcal{C}$ which is not in 3NF. Then there is a "bad" functional dependency $X \to A$ where $A \notin X$, A is nonprime and X does not contain a key. By Theorem 1, the projections $R_1 = R[X \cup A]$ and $R_2 = R[\mathcal{C} - A]$ form a lossless decomposition. If R_1 and R_2 are in 3NF, the decomposition is complete. Otherwise R_1 and/or R_2 are not in 3NF. The technique just described can be applied to the resulting relations that are not 3NF. Each application yields a decomposition into two relations, each of which will contain fewer attributes than the original. Since there are a finite number of attributes, this process must terminate. □

In the relation 0_4 = TOTAL-OFFERINGS [C#, CN, SEC, P#], C# $\to$ CN and C# is not a key. But CN is prime (CN $\subset$ {CN, SEC} which is a key), so this dependency does not violate 3NF. However it still causes the redundancy and anomaly problems mentioned earlier. This can be countered by eliminating the nonprime requirement and thus strengthening the definition of 3NF.

Assume R is a relation, X is a subset of the attributes of R and $A \notin X$.

BCNF: R is in *Boyce Codd Normal Form* if $X \to A$ implies that X contains a key.

Then 0_4 is in 3NF but not BCNF. However it can be decomposed into $0_5 = 0_4$[C#, CN] and $0_6 = 0_4$[C#, SEC, P#]. By Theorem 1, 0_5 and 0_6 is a lossless decomposition. It is also easy to check that 0_5 and 0_6 are in BCNF (see Figure 6). Thus the relations 0_1, 0_3, 0_5 and 0_6 form a lossless BCNF decomposition of TOTAL-OFFERINGS. Clearly, Theorem 2 can be modified to yield the following theorem.

THEOREM 3. *If a relation is not in BCNF, then it is possible to decompose it losslessly into a collection of BCNF relations.* □

The decomposition technique described here is not without problems. In particular, it does not generate a unique decomposition nor will it necessarily generate a decomposition with a minimal number of relations. Consider the BCNF decomposition of OFFERINGS generated above (Fig. 6). Since C# $\to$ CN and C# $\to$ CR, then C# $\to$ {CN, CR}. If this dependency had been used first, the BCNF decomposition would have been COURSES, PROFESSORS and OFFERINGS as in the introduction (see Figs. 3 & 4). This decomposition has only 3 BCNF relations and would be preferable since it would require less space. There are algorithms to insure that a minimal number of 3NF relations are generated [7]. These algorithms involve more theoretical aspects of the set of functional dependencies, some of which are described in Part 4. But there is no known algorithm to use one minimal 3NF decomposition to help generate the remaining minimal decompositions.

In addition, the decomposition may not always yield "natural" relations. Consider the relation R with attributes C# (course number), SEC (section), P# (professor number) and O# (the unique office number for a professor), where we assume professors do not share offices. Then

C#	CN
111	Calculus I
112	Calculus II
243	Linear Algebra
318	Analysis

$O_5 = O_4$ [C#,CN]

C#	SEC	P#
111	1	125
111	2	382
111	3	418
111	4	125
111	5	432
112	1	153
112	2	418
243	1	125
243	2	382
243	3	432
318	1	153

$O_6 = O_4$ [C#,SEC,P#]

FIG. 6

{C#, SEC} → P# while P# → O# and O# → P#. Both of these last two dependencies are "bad." Apply the algorithm using O# → P# to generate R[O#,C#,SEC] and R[O#,P#]. This is a BCNF decomposition. (Note that {C#, SEC} → O# by the transitive property.) But it is not likely that anyone would have a use for a relation that stored the course number, section and office number of the professor teaching the course. General decomposition algorithms do not provide a means to avoid such unnatural decompositions.

Theorems 2 and 3 show that a relation can always be decomposed into either a 3NF or a BCNF collection. But a BCNF decomposition may cause another problem. Consider the relation SCHEDULE with attributes Cl (class, meaning course number and section), S# (student number), and T (time) (Fig. 7). At any time, a student has only one class scheduled but a class can have more than one student. In addition, a class is scheduled for only one time. The functional dependencies for SCHEDULE are {T, S#} → Cl and Cl → T. Thus {T, S#} is a key for the relation. Since T is prime, the dependency Cl → T does not violate 3NF. But it does violate BCNF. The decomposition algorithm can be applied to SCHEDULE to yield the 2 BCNF

C1	S#	T
111-1	12654	1:00
111-1	12785	1:00
111-2	13865	2:00
111-2	17642	2:00
111-2	18725	2:00
243-1	12654	3:00

SCHEDULE

FIG. 7

relations S_1 = SCHEDULE [C1,T] and S_2 = SCHEDULE [C1,S#]. This decomposition is lossless. But the dependency {T,S#} → C1 in SCHEDULE is not a dependency in either S_1 or S_2. Then times could be assigned to classes and inserted in S_1 without violating any dependencies in S_1 while students are being assigned to classes and inserted in S_2, without violating any dependencies in S_2. This could result in scheduling conflicts (e.g., a student being assigned to two classes which meet at the same time). This problem could be handled by adding interrelational dependencies to the database but this is very awkward and is not done in practice. It is apparent that the restricted goal of avoiding redundancies and anomalies in a lossless decomposition does not eliminate all the problems involved with a database. The issue of preserving dependencies will be considered in Section 4.

3. Fourth Normal Form. When some authors examined the semantics of relational databases it became clear that there was often structure beyond that implied by the functional dependencies. For example, suppose we expand the database by adding both the student number (S#) and the possibility of team-teaching. One possible instance is given by the relation REGISTRATION in Fig. 8. Our concept of key is challenged, for now the only nonprime attributes are PN and CR. We still have the functional dependencies P# → PN, CN ↔ C#, C# → CR, and their implications, but we can no longer assert ⟨C#,SEC⟩ → P# or ⟨C#,SEC⟩ → S#. Nonetheless, a class determines its professors and students. For such situations Delobel [**16**], Fagin [**17**], and Zaniolo

C#	CN	SEC	CR	PN	P#	S#
318	Analysis	1	3	Smith	153	17825
318	Analysis	1	3	Jones	125	17825
318	Analysis	1	3	Smith	153	16245
318	Analysis	1	3	Jones	125	16245
318	Analysis	1	3	Smith	153	24692
318	Analysis	1	3	Jones	125	24692

REGISTRATION

FIG. 8

[**29**] introduced the notion of multivalued dependencies: if a relation R has attribute set $\mathcal{Q}$, and if $X, Y \subseteq \mathcal{Q}$, then X *multidetermines* Y, denoted by $X \to\to Y$, if whenever $r_1, r_2 \in R$ and $r_1[X] = r_2[X]$, then there is an $r_3 \in R$ such that $r_3[X \cup Y] = r_1[X \cup Y]$ and $r_3[\mathcal{Q} - Y] = r_2[\mathcal{Q} - Y]$. The definition is symmetric in Y and $\mathcal{Q} - Y$ so $X \to\to Y$ if and only if $X \to\to \mathcal{Q} - Y$. In REGISTRATION, {C#, CN, CR, SEC} $\to\to$ {PN, P#} and {C#, CN, CR, SEC} $\to\to$ S#, for if in a given section of a given course professor A has student 1 and professor B has student 2, then in that same section of that course professor B has student 1 and professor A has student 2.

In such situations one should project onto smaller relations to separate independent facets of the relation. The justification for this is contained in the following proposition. The proof is straightforward and will be omitted.

PROPOSITION 1. *Let R be a relation with attributes $\mathcal{Q}$, and let $X, Y \subset \mathcal{Q}$. Then $X \to\to Y$ if and only if $R = R[X \cup Y] * R[\mathcal{Q} - Y]$.* □

As with functional dependencies, multivalued dependencies have several properties. Among these are:

MD1: $X \to\to Y$ iff $X \to\to \mathcal{Q} - Y$ iff $X \to\to Y - X$.
MD2: If $X \to\to Y$ and $Y \to\to Z$ then $X \to\to Z - Y$.
MD3: If $X \to\to Y$ and $V \subset W$, then $\{W, X\} \to\to \{V, Y\}$.
MD4: If $X \to Y$ then $X \to\to Y$.
MD5: If $X \to\to Z$ and $Y \to Z'$, where $Z' \subset Z$ and $Y \cap Z = \varnothing$, then $X \to Z'$.

Proposition 1 shows multivalued dependencies can be viewed either as semantic constructs or as the extrema of the theory of lossless binary joins. This latter view in particular suggests another normal form, first defined by Fagin [**17**]:

4NF: A relation R is in *fourth normal form* if whenever $X \to\to Y$ holds in R, then $X \to A$ holds for each attribute A of R.

There are two immediate questions: can every relation be decomposed into a lossless family of relations, each of which is in 4NF, and what is the relationship of 4NF to the other normal forms? The former question is handled in the same manner as used to prove that every relation has a 3NF decomposition, using any "bad" dependency to create two subrelations. To answer the latter question we assert that 4NF implies BCNF. To see this, suppose R is in 4NF and suppose R has a functional dependency $X \to Y$. By property MD3, $X \to\to Y$, so by the definition of 4NF, if A is any attribute of R, then $X \to A$. Hence R is in BCNF.

Multivalued dependencies are deceptively similar to functional dependencies, so we should point out differences. For example, there is no decomposition rule, e.g., in REGISTRATION we know {C#, CN, CR, SEC} $\to\to$ {P#, PN}, but it is not true that {C#, CN, CR, SEC} $\to\to$ PN. There is a weak decomposition given in the following proposition.

PROPOSITION 2. *Let R be a relation with attributes $\mathcal{Q}$, and suppose $X \to\to Y$ in R. Let $Y' \subset Y$ and $Z \subset \mathcal{Q} - (X \cup Y)$. Then $X \to\to Y'$ in $R' = R[X \cup Y' \cup Z]$.*

Proof. Let $r_1, r_2 \in R'$ be such that $r_1[X] = r_2[X]$, and let $s_1, s_2 \in R$ be such that $s_i[X \cup Y' \cup Z] = r_i$ for $i = 1, 2$. Then there is an $s_3 \in R$ such that $s_3[Y] = s_1[Y]$ and $s_3[\mathcal{Q} - Y] = s_2[\mathcal{Q} - Y]$. If $r_3 = s_3[X \cup Y' \cup Z]$, then $r_3[Y'] = r_1[Y']$, $r_3[X \cup Z] = r_2[X \cup Z]$. □

The loss of decomposition (property FD5) leads to difficulties because we must be more careful when trying to reduce a set of multivalued dependencies to a simpler equivalent set. More troublesome problems occur when we try to combine projections and dependencies. For functional dependencies, $X \to Y$ in $R[X \cup Y]$ if and only if $X \to Y$ in R, but $X \to\to Y$ is always true in $R[X \cup Y]$. Even worse, Fagin [**17**, Table V] gives an example of a relation S with attributes W, X, Y, Z such that $X \to\to Y$ in $S[X \cup Y \cup Z]$, but there are no nontrivial multivalued

dependencies in S. In this case S is said to have an *embedded multivalued dependency*, denoted $X \rightarrow\rightarrow Y|Z$. The proof of the existence of a 4NF decomposition for every relation is still valid in the presence of embedded multivalued dependencies, but their presence complicates the selection of a good decomposition. They also lead to problems in axiomatizing dependencies. Fortunately, Fagin asserts that such embedded dependencies rarely occur in practice [**17**].

As an example of a 4NF decomposition, we will decompose the relation REGISTRATION in Fig. 7. We first remove the simple dependencies C# $\rightarrow$ CR, C# $\leftrightarrow$ CN, and P# $\rightarrow$ PN by forming the relations COURSES, with attributes {C#, CN, CR}, PROFESSORS, with attributes {P#, PN}, and a temporary relation $R1$, with attributes {C#, SEC, P#, S#}. Notice that REGISTRATION = COURSES $*$ $R1$ $*$ PROFESSORS, and both COURSES and PROFESSORS are in 4NF. To decompose R_1 further, we note we have {C#, SEC} $\rightarrow\rightarrow$ {P#, PN} in REGISTRATION, so by Proposition 1 we have {C#, SEC} $\rightarrow\rightarrow$ P# in $R1$. We form the relations TEACHING-ASSIGNMENTS, with attributes {C#, SEC, P#}, and COURSE-WORK, with attributes {C#, SEC, S#}. Each of these subrelations is in 4NF, and their join is $R1$, so we have produced the desired decomposition. Figure 9 shows the resulting values.

C#	SEC	P#
318	1	153
318	1	125

TEACHING - ASSIGNMENTS = REGISTRATION [C#,SEC,P#]

C#	SEC	S#
318	1	17825
318	1	16245
318	1	24692

COURSE - WORK = REGISTRATION [C#,SEC,S#]

FIG. 9

In our example the 4NF version is superior to the less decomposed versions. However, since 4NF implies BCNF, there are examples, mentioned earlier, where the 4NF may force decompositions which do not preserve functional dependencies. To avoid such problems, there is also a *weak fourth normal form* (W4NF): a relation R with attributes $\mathcal{Q}$ is in W4NF if it is in 3NF and whenever $X \rightarrow\rightarrow Y$, then $\mathcal{Q} = X \cup Y$. Clearly 4NF implies W4NF implies 3NF. As before, a simple cardinality argument shows that every relation has a W4NF decomposition. 4NF is more widely discussed than W4NF primarily because it was introduced first, but some of the early semantic arguments may be viewed as arguing more for W4NF than for 4NF.

One can go beyond multivalued dependencies and define join dependency [**1**], [**26**]. Given a relation R with attributes $\mathcal{Q}$ and a collection $\{X_i\}_{i=1}^n$, where each $X_i \subseteq \mathcal{Q}$, then *$R$ jointly depends on* the $\{X_i\}$ if $R = R[X_1] * \cdots * R[X_n]$. Join dependencies seem to have arisen purely from theoretical considerations, pushing the theory of lossless joins to their limits. Fagin [**19**] uses them to define a join normal form (JNF) in which there are no nontrivial joint dependencies. However,

we do not pursue this because in realistic examples join dependencies never seem to arise except when they are implied by simpler dependencies. Further dependencies, again more motivated by theory than by practice, are discussed in [22]. Finally, Fagin [20] has recently introduced a further normalization, domain-key normal form (DKNF), which is motivated by the observation that all current database systems routinely will check that each data value entered is in the correct domain and that all keys are distinct. While it is interesting, we do not pursue this normal form primarily because we are not convinced that database design should be reduced to the lowest common denominator of database systems, especially at so early a stage in the history of databases.

4. Dependencies. This section looks a little more deeply into the theory of dependencies. Assume F is the given set of functional dependencies on a set of attributes $\mathcal{Q}$. The *closure* of F, denoted F^+, is the smallest superset of F which is closed under axions FD1—FD3 (projectivity, additivity and transitivity—see Part 2). That is, any instance of a relation R with attributes $\mathcal{Q}$ that satisfies the functional dependencies in F also satisfies all the dependencies in F^+. Armstrong [2] proved that these axioms were complete in the sense that there exists an instance of a relation R which has attributes $\mathcal{Q}$ and which satisfies the functional dependencies in F^+ and no others. (Armstrong's proof actually involved a set of axioms equivalent to FD1—FD3.) Note that F^+ can be quite large even if F is small. For example, if $A \to B_1$, $A \to B_2, \dots, A \to B_n \in F$, then $A \to Y \in F^+$ for Y any subset of $\{B_1, \dots, B_n\}$. Since $\{B_1, \dots, B_n\}$ has 2^n subsets, computing the closure can be a time-consuming process. However, there exist efficient algorithms to determine whether a particular functional dependency is in F^+ [3].

A set of dependencies G is a *cover* for F if $G^+ = F^+$. G is a *minimal* cover if the following properties hold:

C1. The right side of every dependency in G consists of a single attribute.
C2. G has no redundant dependencies, where a dependency $X \to A$ is redundant if $G^+ = (G - \{X \to A\})^+$.
C3. The left side of every dependency in G has no redundant attributes, where Z is a redundant attribute of $Z \cup X \to A$ if $X \to A$ and $Z \cap X = \phi$.

The existence of a cover satisfying C1 for any set of functional dependencies F is an easy consequence of the additive (FD2) and decomposition (FD5) axioms. To generate a cover satisfying C2, take a cover of F which satisfies C1 and consider each dependency in some order. If a dependency is redundant, remove it from the set. To also satisfy C3, check each dependency and each attribute in its left side in some order. If an attribute is redundant, then remove it from the dependency. This will yield a cover satisfying C3, and thus every set of dependencies has a minimal cover. A minimal cover is not necessarily unique since considering the dependencies in a different order may yield different minimal covers.

The theory of functional dependencies (closure and covers) has been related to other mathematical areas. Delobel and Casey [15] developed an equivalence between functional dependencies and Boolean functions such that axioms applied to one are equivalent to analogous axioms applied to the other. Fagin [18] developed an equivalence between functional dependencies and implication statements of propositional logic. In particular, he showed that the closure of a set of functional dependencies is equivalent to the closure under analogous operations of the equivalent set of implication statements. Readers wishing to pursue such connections are encouraged to read [21].

It is now possible to formalize what it means to preserve dependencies. Assume R_1 is a projection of the relation R. The set F_1 of projected dependencies for R_1 is the set of all dependencies, $X \to Y$, in F^+ such that all the attributes in $X \cup Y$ are attributes of R_1. The decomposition of R into the relations $R_1, \dots, R_n$ is *dependency preserving* if $F^+ = (F_1 \cup \cdots \cup F_n)^+$ where F_i is the set of projected dependencies for R_i, $i = 1, \dots, n$.

As the relation SCHEDULE (Fig. 7) shows, it is not always possible to find a lossless BCNF

decomposition which also preserves dependencies. It is possible to determine whether such a decomposition exists. Form all possible dependency preserving decompositions and then check each one to see if it is BCNF. However, this algorithm requires so much time it is unusable in practice, and so far no suitable algorithm has been developed. In fact, [**4**] showed that this problem is NP-hard, which means it is not likely a good algorithm can be found.

On the other hand, there always exists a lossless 3NF decomposition which preserves dependencies. This is accomplished by the following algorithm:

Decomposition Algorithm. Assume R is a relation and F is a minimal cover of the functional dependencies for R.

(1) If an attribute does not appear on the left or right side of any dependency in F, then it can form a relation on its own and be deleted from R.
(2) Partition F into groups such that all dependencies in a group have the same set of attributes on the left.
(3) Combine the dependencies in each group into a single dependency using the additive axiom (FD2).
(4) For each of the resulting dependencies, $X \to Y$, form the projection $R[X \cup Y]$.

The resulting decomposition is clearly dependency preserving. Since a minimal cover has no redundant attributes on the left side of any dependency, the set of attributes X is a key for $R[X \cup Y]$. Thus it is in 3NF. As it stands, such a decomposition may be lossy. But if the projection $R[X]$, where X is a key for R, is included in the decomposition, then it is lossless. This can be checked by applying the lossless join algorithm of Aho, Beeri, and Ullman [**1**].

The relation TOTAL-OFFERINGS (Fig. 1) can be decomposed using this algorithm. The initial set F (see Part 2) of functional dependencies is in fact minimal. The decomposition would consist of the following projections: TOTAL OFFERINGS [C#, CN, CR], TOTAL-OFFERINGS [CN, C#], TOTAL-OFFERINGS [P#, PN], TOTAL-OFFERINGS [C#, SEC, P#], and TOTAL-OFFERINGS [C#, SEC] since C#, SEC is a key for TOTAL-OFFERINGS. Since this set F has other minimal covers, the decomposition is not unique. Nor is it minimal. But the collection can be reduced by eliminating projections whose attributes are contained in another projection. In this example, the remaining projections would be TOTAL-OFFERINGS [C#, CN, CR], TOTAL-OFFERINGS [P#, PN], TOTAL-OFFERINGS [C#, SEC, P#]. In general, the decomposition algorithm does not yield unique nor minimal decompositions.

Applying this algorithm to the relation SCHEDULE [C1, S#, T] yields the projections SCHEDULE [C1, T] and SCHEDULE [C1, S#, T]. After simplifying, the only remaining relation is the original one itself which is 3NF but not BCNF.

So far, none of this theory takes into account other possible connections between attributes. For example, suppose X and Y are attributes of a relation R for which $r_1, r_2 \in R$, $r_1 \neq r_2$, $r_1[X] = r_2[X]$ implies $r_1[Y] \neq r_2[Y]$. To exploit this information, you want the decomposition to keep X and Y together. Bernstein [**7**] suggested that in this situation we introduce a dummy attribute θ and include the functional dependency, $X, Y \to \theta$. After the decomposition algorithm has been applied, the dummy attribute is dropped. (Ling, Tompa, and Kameda [**23**] also study this problem.) Apparently such connections rarely occur in practice, or if they do occur, they are ignored.

The concept of preserving multivalued dependencies has been studied, but not as extensively. Let D be the given set of functional and multivalued dependencies for a relation R. The closure of D, denoted by $D+$, is the smallest superset of D which is closed under axioms FD1–FD3 and MD1–MD5. Beeri, Fagin and Howard [**6**] showed that these axioms are complete. The computation of $D+$ may again require an amount of time which is an exponential function of the number of attributes and given dependencies. However, there is an efficient algorithm to determine whether a particular multivalued dependency is in $D+$ [**3**]. As shown earlier, there always exists a

decomposition of a relation into 4NF. However, since 4NF implies BCNF, it is not always possible to generate a 4NF decomposition that preserves functional dependencies. The existence of embedded dependencies also complicates the decomposition process.

5. Conclusion. In this paper we have roughly followed the historical development of the design of relational databases, which essentially started with the early papers of Codd [**9**], [**10**], [**11**]. Codd [**11**] listed the advantages of the relational database as being:

1) a high degree of data independence;
2) a community view of the data of spartan simplicity so that a wide variety of users can interact with a common model;
3) a simplification of the job of the database administrator;
4) a theoretical foundation for database management;
5) a merging of fact retrieval and file management fields;
6) a lifting of databased application programming to a level where sets are treated as operands.

There then followed a period of discovering various decompositions and normal forms [**17**], [**19**], [**20**], [**22**]. Normal forms are motivated by the desire to compress the amount of data being stored, without losing the ability to regenerate all the original data. Unfortunately, as demonstrated in Part 2, this can run counter to attempts to preserve other information, such as functional dependencies, and if applied blindly may lead to relations being stored which have no natural interpretation. The role of the semantics of the database, i.e., what the data and relations mean, as opposed to only their syntax, i.e., the ways they can be combined, is becoming increasingly important. We expect that considerable effort will be spent trying to combine the space efficiencies of the normal forms with the demands that the relations be meaningful and that no information be lost. One proposal concerning semantics is the notion of a well-defined relation offered by Smith and Smith [**27**]. While we cannot easily give a precise logical definition of well-defined, the major requirement of such a relation is that one be able to assign to it and all of its attributes names which are natural. The relation in Part 2 with attributes course number, section, and office number of the professor teaching the course is not well-defined since there is no natural name for it. Other semantic concerns are discussed by Codd [**12**].

There are a large number of closely related, mathematically oriented problems which we have only hinted at. For example, in Part 4 we mentioned that some decomposition questions have been shown to be NP-hard. One such question is the existence of a BCNF decomposition that preserves dependencies [**4**]. This is not necessarily as bad as it seems, since in practice a person can usually decide this problem quite quickly. As the determination of such decompositions is increasingly being performed by computers, it does indicate that attention needs to be focused on finding computationally efficient heuristics to help find such a decomposition. In particular, some effort has been spent to find subclasses of relations for which a desired decomposition can be found quickly.

One closely related area which was not touched in this paper is the problem of actually accessing the data stored in a relational database. For example, to answer the query "What courses are taught by both Professor Jones and Professor Thomas?", how should one combine the relations in Fig. 4 to answer this question most efficiently? Of course, the first problem is to translate English queries into mathematical statements, but once this has been done, one sees that a given problem can be solved in many ways, some of which are far more efficient than others. A great deal of work has been done on this problem (see, e.g., Ullman, Chap. 6 [**28**]), but these considerations go beyond the aims of this paper.

One final, important issue concerns the practicality of using relations as the foundation of an actual database management system (DBMS). In his Turing Award acceptance lecture, delivered in November 1981 [**13**], Codd stated

> There has been no shortage of skepticism concerning the practicality of the relational approach to database management.... Instead of welcoming a theoretical foundation as providing soundness, the attitude seems to be: if it's theoretical, it cannot be practical. The absence of a theoretical foundation for almost all nonrelational DBMS is the prime cause of their ungepotchket quality. (This is a Yiddish word, one of whose meanings is patched up.)

Despite early resistance, relational DBMS seem to be gaining acceptance. Two commercially available systems are Tandem Computer Corporation's ENCOMPASS and IBM's System R, and early reports on their performance are positive [**13**].

References

1. A. V. Aho, C. Beeri, and J. D. Ullman, The theory of joins in relational databases, ACM Transaction on Database Systems, 4 (1978) 297–314.

2. W. W. Armstrong, Dependency structures of data base relationships, Information Processing, 1974, North Holland, Amsterdam, pp. 580–583.

3. C. Beeri, On the membership problem for multivalued dependencies in relational databases, TR229, Dept. of EECS, Princeton University, Princeton, NJ, 1977.

4. C. Beeri and P. A. Bernstein, Computational problems related to the design of normal form relation schemes, ACM Transactions on Database Systems, 4 (1979) 30–59.

5. C. Beeri, P. A. Bernstein, and N. Goodman, A sophisticate's introduction to database normalization theory, Proceedings ACM International Conference on Very Large Databases, 1978, pp. 113–124.

6. C. Beeri, R. Fagin, and J. H. Howard, A complete axiomatization for functional and multi-valued dependencies, ACM/SIGMOD International Symposium on Management of Data (1977) 47–61.

7. P. A. Bernstein, Synthesizing third normal form relations from functional dependencies, ACM Transactions on Database Systems, 1 (1976) 277–298.

8. A. F. Cardenas, Data Base Management System, Allyn and Bacon, Boston, Mass., 1979.

9. E. F. Codd, A relational model of data for large shared data banks, Comm. ACM 13 (1970) 377–387.

10. ______, Further normalization of the data base relational model, Courant Computer Science Symposia 6: Data Base Systems, Prentice-Hall, Englewood Cliffs, NJ, 1971, pp. 65–98.

11. ______, Recent investigations in relational data base systems, Information Processing, 1974, North Holland, Amsterdam, 1974, pp. 1017–1021.

12. ______, Extending the data base relational model, ACM Transactions on Database Systems, 4 (1979) 397–434.

13. ______, Relational database: a practical foundation for productivity, Communications ACM, 25 (1982) 109–117.

14. C. J. Date, An Introduction to Database Systems, Addison-Wesley, Reading, Mass., 1977.

15. C. Delobel and R. G. Casey, Decomposition of a data base and the theory of Boolean switching functions, IBM Journal of Research Development (1973) 374–485.

16. C. Delobel, Normalization and hierarchical dependencies in the relational data model, ACM Transactions on Database Systems, 3 (1978) 201–222.

17. R. Fagin, Multivalued dependencies and a new normal form for relational databases, ACM Transactions on Database Systems, 2 (1977) 262–278.

18. ______, Functional dependencies in a relational database and propositional logic, IBM Journal of Research Development (1977) 534–544.

19. ______, Normal forms and relational data base operators, ACM/SIGMOD International Symposium on Management of Data (1979) 153–160.

20. ______, A normal form for relational databases that is based on domains and keys, ACM Transactions on Database Systems, 6 (1981) 397–415.

21. H. Gallaire and J. Minker, Logic and Data Bases, Plenum Press, NY (1978).

22. H. Gallaire, J. Minker, and J. Nicolas, Advances in Data Base Theory, vol. 1, Plenum Press, NY (1981).

23. T. K. Ling, F. W. Tomja, and T. Kameda, An improved normal form for relational databases, ACM Transactions on Database Systems, 6 (1981) 329–346.

24. J. Martin, Computer Data-Base Organization, Prentice-Hall, Englewood Cliffs, NJ, 1977.

25. J. Rissanen, Independent components of relations, ACM Transactions on Database Systems, 2 (1977) 317–325.

26. ______, Theory of relations for databases—A tutorial survey, Proceedings 7th Symposium on Mathematical Foundations of Computer Science, Lecture Notes in Computer Science, No. 64, Springer Verlag, NY, 1978.

27. J. M. Smith and D. C. P. Smith, Database abstractions: aggregation, Communications ACM, 20 (1977) 405–413.
28. J. D. Ullman, Principles of Database Systems, Computer Science Press, Potomac, Maryland, 1980.
29. C. Zaniolo, Analysis and design of relational schemata for database systems, Ph.D. Thesis, Tech. Rep. UCLA-Eng-7669, Univ. of Calif., Los Angeles, 1976.
30. New computerized retrieval service, Notices Amer. Math. Soc., 28 (1981) 449–451.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A MISCELLANY OF ERDŐS PROBLEMS

Paul Erdős continues to pour out problems. Here are several that he has handed to me in recent months.

If $1 \leqslant a_1 < a_2 < \cdots < a_k \leqslant n$, how large can k be if no $a_j - a_i$ divides a_j for $1 \leqslant i < j \leqslant k$?

If $k > \lceil n/2 \rceil$, the least integer not less than $n/2$, then $a_{i+1} - a_i = 1$ for some i. On the other hand, by taking $a_i = 2i - 1$, we see that this is best possible. If, however, we ignore the divisor $a_j - a_i = 1$, then we can take the odd powers of 2 along with the odd numbers, i.e., $k \geqslant \lfloor n/2 \rfloor + \lfloor (\operatorname{lb} n)/2 \rfloor$ where lb is the binary (base 2) logarithm. More generally we can ask:

Is it true that if $k > (\frac{1}{2} + \varepsilon)n$, then there is a "large" difference $a_j - a_i$ which divides a_j?

To make this more precise, assume

$$(a_j - a_i) \nmid a_j \text{ whenever } a_j - a_i \geqslant t \tag{1}$$

and write max $k = F(n; t)$ for all sequences satisfying (1). Is it true for every t that

$$¿ \lim_{n \to \infty} F(n; t)/n = 1/2 ?$$

We have seen that $F(n; 1) = \lceil n/2 \rceil$ and that $F(n; 2) > n/2 + c \ln n$. Can you show $F(n; 2) < (1 + \varepsilon)n/2$? What about larger values of t?

Erdős writes that in a recent letter to him, I. Ruzsa proved that $F(n, t) < n/2 + c_t n/\ln n$. The lower bound hasn't been improved.

A Problem of Erdős and Marcel Erné

Let $G(n)$ be a graph on n vertices. Denote by $f(G(n))$ the number of complete graphs contained in $G(n)$, and by $F(n)$ the number of possible values of $f(G(n))$. For example $F(1) = 1$ and, for $n = 2$, $f(G) = 2$ or 3, so $F(2) = 2$. For $n = 3$, $f(G)$ is 3, 4, 5 or 7, so $F(3) = 4$. If my doodlings are correct, $F(4) = 8$ and $F(5) = 16$. [Another example of the strong law of small numbers?] But Erdős says that $F(n) = o(2^n)$ is easy to show and that we almost certainly can prove that

$$\lim_{n \to \infty} F(n)^{1/n} = 2.$$

27. J. M. Smith and D. C. P. Smith, Database abstractions: aggregation, Communications ACM, 20 (1977) 405–413.
28. J. D. Ullman, Principles of Database Systems, Computer Science Press, Potomac, Maryland, 1980.
29. C. Zaniolo, Analysis and design of relational schemata for database systems, Ph.D. Thesis, Tech. Rep. UCLA-Eng-7669, Univ. of Calif., Los Angeles, 1976.
30. New computerized retrieval service, Notices Amer. Math. Soc., 28 (1981) 449–451.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A MISCELLANY OF ERDŐS PROBLEMS

Paul Erdős continues to pour out problems. Here are several that he has handed to me in recent months.

If $1 \leqslant a_1 < a_2 < \cdots < a_k \leqslant n$, how large can k be if no $a_j - a_i$ divides a_j for $1 \leqslant i < j \leqslant k$?

If $k > \lceil n/2 \rceil$, the least integer not less than $n/2$, then $a_{i+1} - a_i = 1$ for some i. On the other hand, by taking $a_i = 2i - 1$, we see that this is best possible. If, however, we ignore the divisor $a_j - a_i = 1$, then we can take the odd powers of 2 along with the odd numbers, i.e., $k \geqslant \lfloor n/2 \rfloor + \lfloor (\text{lb}\, n)/2 \rfloor$ where lb is the binary (base 2) logarithm. More generally we can ask:

Is it true that if $k > (\frac{1}{2} + \varepsilon)n$, then there is a "large" difference $a_j - a_i$ which divides a_j?

To make this more precise, assume

$$(a_j - a_i) \nmid a_j \text{ whenever } a_j - a_i \geqslant t \tag{1}$$

and write max $k = F(n; t)$ for all sequences satisfying (1). Is it true for every t that

$$¿ \lim_{n \to \infty} F(n; t)/n = 1/2 ?$$

We have seen that $F(n; 1) = \lceil n/2 \rceil$ and that $F(n; 2) > n/2 + c \ln n$. Can you show $F(n; 2) < (1 + \varepsilon)n/2$? What about larger values of t?

Erdős writes that in a recent letter to him, I. Ruzsa proved that $F(n, t) < n/2 + c_t n/\ln n$. The lower bound hasn't been improved.

A Problem of Erdős and Marcel Erné

Let $G(n)$ be a graph on n vertices. Denote by $f(G(n))$ the number of complete graphs contained in $G(n)$, and by $F(n)$ the number of possible values of $f(G(n))$. For example $F(1) = 1$ and, for $n = 2$, $f(G) = 2$ or 3, so $F(2) = 2$. For $n = 3$, $f(G)$ is 3, 4, 5 or 7, so $F(3) = 4$. If my doodlings are correct, $F(4) = 8$ and $F(5) = 16$. [Another example of the strong law of small numbers?] But Erdős says that $F(n) = o(2^n)$ is easy to show and that we almost certainly can prove that

$$\lim_{n \to \infty} F(n)^{1/n} = 2.$$

Similar questions can be asked about **cliques** (*maximal* complete subgraphs) and other types of subgraph (e.g., stars, circuits, trees, bipartite graphs).

Joel Spencer improved on earlier results of Moon and Moser and of Erdős and showed that there is a graph $G(n)$ which contains $n - \mathrm{lb}\, n + O(1)$ cliques of distinct size and it is easy to see that this result is best possible. What is the corresponding result for hypergraphs (graphs in which the "edges" are subsets of the vertices whose cardinality may be greater than 2)? If $L(n)$ is the number of times we have to iterate the logarithm to reach a number less than e, then Erdős has shown that there is a 3-graph on n vertices (triple system on n elements) which contains $n - L(n)$ cliques of distinct size (a clique is here a maximal set of vertices all of whose subsets of size 3 are edges). Can it be shown that there is no such 3-graph (triple system) with $n - c$ cliques of distinct size, for some constant c?

A Problem of Erdős and William T. Trotter

Let S be a set of cardinality n and $\{A_k\}$ be a family of subsets of S with no A_k containing any other. Suppose that for each t, if there is an A_k of size t, then there are at least r different A_k of that size; i.e., any size which occurs must occur at least r times. What is the largest number of subsets of different size?

If $r = 1$ and $n > 3$, there are always $n - 2$ sets of distinct size, but you can't find $n - 1$. If $r > 1$ and $n > n_0(r)$, there are always $n - 3$ sets of distinct size, but there aren't $n - 2$. We have no satisfactory estimate of $n_0(r)$.

The content of the previous two paragraphs will be a small subset of a forthcoming paper of Erdős, Szemerédi and Trotter.

A Problem of Erdős and David E. Daykin

Let $F = \{A_1, A_2, \ldots, A_m\}$ be a family of m subsets of a set S of size $2n$. Consider the graph $G(F)$ whose vertices are the A_i, two vertices being joined just if the corresponding A_i are **comparable**, i.e., just if one contains the other. An old conjecture of Erdős is that if $m = (2 - \varepsilon)2^n$ and $n > n_0$, then $G(F)$ has fewer than 2^{2n} edges. It is easy to see that this is false if $\varepsilon = 0$: split S into two parts, B and C, each of size n, and for A_i, $1 \leqslant i \leqslant 2^n$, take all the subsets of B, and for A_i, $2^n < i \leqslant 2^{n+1}$, take the union of B with every subset of C. Then $A_i \subset A_j$ whenever $1 \leqslant i \leqslant 2^n < j \leqslant 2^{n+1}$, giving 2^{2n} edges in $G(F)$.

Daykin and P. Frankl proved that if $G(F)$ has $(1 + o(1))\binom{m}{2}$ edges (i.e., if the graph is complete apart from $o(m^2)$ edges), then $m^{1/n} \to 1$ as $n \to \infty$. Daykin and Erdős conjecture that if $G(F)$ has cm^2 edges for some $c > 0$ and independent of n, then $m < (2 + o(1))^n$. Perhaps there is a C, depending on c, so that $m < C \cdot 2^n$ for sufficiently large n. If true, this conjecture is best possible. Also perhaps to every $\varepsilon > 0$ there corresponds an η such that if $G(F)$ has more than $m^{2-\eta}$ edges, then $m < (2 + \varepsilon)^n$. Perhaps an Erdős-Stone type result holds: if $G(F)$ has more than $(1 + \varepsilon)\frac{m^2}{2}\left(1 - \frac{1}{r}\right)$ edges, then $m < 2^{n(1+o(1))/(r+1)}$. Many further related questions can be asked.

Three Games Played by Coloring Edges of Graphs

The first game is played on the complete graph, K_n, on n vertices, each pair of vertices being joined by an edge. Left starts by coloring an edge Lilac, and Right replies by coloring another edge Red. They continue to color new edges alternately until all the $\binom{n}{2}$ edges are colored. Left wins if he can produce a larger Lilac clique (complete subgraph, K_r) than any of Right's Red cliques. If Right produces a clique as large or larger than any of Left's, then she wins. Erdős conjectures that, for $n > 2$, the second player, Right, wins.

In the second game, Right colors two edges after each time Left colors one. For $n > 3$, can Right always produce a larger clique? She can for $n = 4$ or 5.

In the third game they color alternately, as in the first game, but Left wins if he can produce a vertex with larger Lilac valence than any Red valence, i.e., if he can make more Lilac edges at a vertex than Right can make Red edges at any vertex. If the maximum Lilac and Red valences are equal, Right wins. Who does win?

A Conjecture of Marian Deaconescu

Let $\Pi = p_1 p_2 \cdots p_n$ be the product of the first n primes. Then

¿ $\Pi + p$ is prime for at least one prime p, $p_n < p < \Pi$?

Erdős agrees with this conjecture and expects that the least such p is much smaller than Π; in fact that $p < n^c$ for some constant c. Deaconescu has verified the conjecture for $n \leqslant 1000$.

A PENTAD OF POINTED PROBLEMS

Quite often the MONTHLY gets problems which are too unsolved to go into the elementary or advanced sections, even with stars on, while they may be too brief to stand alone in this section. Here are five examples.

Some Problems in Geometric Probability

Paul R. Chernoff, University of California, Berkeley, CA 94720

Choose n points $P_1, P_2, \ldots, P_n$ at random from the unit disk in R^2. What is the probability that the polygon $P_1P_2 \ldots P_nP_1$ is convex? What is the probability that it's a Jordan curve? If the points are chosen from the unit ball in R^3, what's the probability that the polygon is knotted?

I have no idea how to solve these, I believe that Dvoretzky has done some possibly related work on Brownian motion trajectories. I think that the first question may be easy, but that the last one is hard.

Find an Invertible Linear Combination

Robert Hartwig, North Carolina State University at Raleigh, NC 27650

If Q_1, Q_2, Q_3, Q_4 are square matrices of order n with elements in a field F and the matrix $\begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix}$ of order $2n$ is invertible, are there scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in F so that $U = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4$ is invertible?

What about the same problem with "field" replaced by "division ring" or "the ring of integers"? A necessary condition for matrices over a commutative ring with unity, is that the **splitting condition** holds; that $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ similar to $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$ implies A similar to B. For, if A, B are n by n matrices, then

$$\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$$

implies $AQ_i = Q_iB$ for each i, and hence since R is commutative, $AU = UB$ so that an affirmative answer to the question implies the splitting condition.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, T(13: 1), L*. The Complete Problem Solver. John R. Hayes. Franklin Inst Pr, 1981, v + 255 pp, $19.50. [ISBN: 0-89168-028-4] Discussion of skills for problem solving (mathematical and otherwise) and the psychology of problem solving. Very well-organized, clearly written, with references and a modest number of exercises. JRG

General, S(13-16), L.** Mathematical Brain Benders: 2nd Miscellany of Puzzles. Stephen Barr. Dover Pub, 1982, 224 pp, $4 (P). [ISBN: 0-486-24260-9] An unabridged and corrected republication of the 1969 work Second Miscellany of Puzzles--Mathematical and Otherwise. Sixty-five well-illustrated, witty, original challenges, mostly geometrical, plus fifty "tricky or overliteral...more or less from everyday life" short puzzles. A truly different kind of collection. Watch out for the unexpected. Highly recommended. The low price is unbeatable. JK

General, T(13: 1), S*, L*. Basic Algebra and Geometry for Scientists and Engineers. A.J. Ellis. Wiley, 1982, xi + 187 pp, $24.95. [ISBN: 0-471-10174-5] Title misleads. Should read "...for students in the sciences and engineering." Linear systems, determinants and matrices. Vectors in two and three dimensions, including complex numbers. Conics and quadric surfaces. Polynomials and quadratic forms in two and three variables. Straightforward. Concise. Clear. Worked-out examples. Applications. Exercises. Answers to many. Useful for students about to take--or currently taking--a calculus course. Handy reference. JK

General, S(18), P, L. A Panorama of Pure Mathematics As Seen by N. Bourbaki. Jean Dieudonné. Trans: I.G. Macdonald. Academic Pr, 1982, x + 289 pp, $29.50. [ISBN: 0-12-215560-2] A survey of those areas of mathematics which have been exposed in the 30+ years of the Bourbaki Seminar. Included are topology (algebraic, differential, but not general), geometry (differential, analytic, algebraic), algebra (homological, number theory, group theory, commutative, but not general), analysis (harmonic, ordinary differential equations, partial differential equations, ergodic theory, but not integration and classical analysis). A remarkable book, one which very few mathematicians could have written. SG

General, L. Mathematics for the General Reader. E.C. Titchmarsh. Dover, 1981, 197 pp, $3 (P). [ISBN: 0-486-24172-6] Lucid introduction for laymen to the ideas of mathematics, from counting to calculus. Unabridged republication of the 1959 Doubleday edition. GHM

General, P. Emil Artin: Collected Papers. Ed: Serge Lang, John T. Tate. Springer-Verlag, 1965, xvi + 560 pp, $30. [ISBN: 0-387-90686-x] All the papers (but not the books or lecture notes), each in the language in which it was published. JD-B

General, T(13: 1), S, L. A Number for Your Thoughts. Stephen P. Richards (Box 501, New Providence, NJ 07974), 1982, 207 pp, $7.95 (P). [ISBN: 0-9608224-0-2] Twenty-one easy to read, informative articles about numbers (primes, rationals, irrationals, complex, transfinite) and number curiosities, properties, and conjectures. These expository essays require no mathematical background, and could become the focus for a continuing education course in mathematics. No exercises. LCL

General, P. The Study of Time IV. Ed: J.T. Fraser, N. Lawrence, D. Park. Springer-Verlag, 1981, xxv + 286 pp, $39. [ISBN: 0-387-90594-4] Papers presented at the fourth (1979) conference of the International Society for the Study of Time. Concludes with a major bibliography on the literature of time, 1900-1980, and a cumulative index of the proceedings of all previous conferences. LAS

General, P. Transactions of the Moscow Mathematical Society, 1982, Issue 1. Ed: Ben Silver. AMS, 1982, v + 298 pp, $91 (P). Translation of Volume 41 (1980).

Elementary, T(13). Intermediate Algebra. John G. Michaels, Norman J. Bloch. McGraw-Hill, 1982, xiii + 417 pp, $15.95. [ISBN: 0-07-041820-9] Includes graphing of conic sections and rational

functions, introduces logarithms. No synthetic division, binomial theorem, or probability. AWR

Precalculus, T*(13: 1). Precalculus Mathematics, A Short Course. Russell E. Thompson. U Pr of America, 1982, viii + 189 pp, $10 (P). [ISBN: 0-819-2634-9] Designed to provide the essentials of pre-calculus in one semester. Brief. Unpretentious. Without frills. Surprisingly thorough. Prepared from typewritten pages. Plastic binder. Exercises. Answers. Appendices. Four-place tables of common logs and trigonometric functions. JK

Education, S, P, L.** Mindstorms: Children, Computers, and Powerful Ideas. Seymour Papert. Basic Books, 1980, viii + 230 pp, $6.68 (P). [ISBN: 0-465-04629-0] A celebration of dynamic approaches to geometry via LOGO and Turtle Geometry--Papert's computer tools designed for kids. Papert's thesis is that young children can explore subtle geometric, algorithmic and computational ideas if only they are placed in an environment with appropriate, enjoyable, playful tools. LAS

Education, P. So You're a Mathematics Supervisor. Ross Taylor. NCTM, 1982, 20 pp, $2.50 (P). [ISBN: 0-87353-199-X] Helpful hints for management in any field. While many suggestions appear to be common knowledge (e.g., let a secretary screen your calls), the book provides reinforcement for experienced supervisors and a glimpse of necessary supervisory techniques for beginners. MW

Education, P. Mathematical Problem Solving: Issues in Research. Ed: Frank K. Lester, Joe Garofalo. Franklin Inst Pr, 1982, xii + 139 pp, $14.50 (P). [ISBN: 0-89168-049-7] Revisions of papers presented at a 1981 conference by researchers active in the field of mathematical problem solving. Consistent theme is the need to draw upon the work of cognitive psychologists. Much attention to stages of problem solving, content and organization of knowledge, and information processing approaches. Information on recent activities and ideas for new research. MW

Education, P. Computers in the Classroom. Ed: Henry S. Kepner, Jr. NEA, 1982, 158 pp, $7.95 (P). [ISBN: 0-8106-1825-7] An introductory survey--already dated--of the ways computers are used in the classroom, ranging from elementary classes and computer literacy to interesting applications in high school mathematics. LAS

History, S(14-18), P, L.** Aristarchus of Samos: The Ancient Copernicus. Sir Thomas Heath. Dover Pub, 1981, viii + 425 pp, $7 (P). [ISBN: 0-486-24188-2] Republication of Heath's classic work of 1913. History of Greek astronomy from Thales through Aristarchus, originator of the heliocentric hypothesis. Includes translation and discussion of Aristarchus only extant work "On the Sizes and Distances of the Sun and Moon." JRG

History, S(16-17), P. Histoire et Préhistoire de L'Analyse des Données. J.-P. Benzécri. Dunod, 1982, vii + 159 pp, 55 FF (P). [ISBN: 2-04-015467-1] A brief history of statistics by a statistician, requiring of the reader considerable knowledge of the field. JD-B

History, P, L. Writings of Charles S. Peirce: A Chronological Edition, Volume I, 1857-1866. Charles S. Peirce. Indiana U Pr, 1982, xxxv + 698 pp, $32.50. [ISBN: 0-253-37201-1] Meticulously documented selection from the manuscripts of this major American philosopher and scientist (1839-1914). A twenty volume series is projected, drawing from Peirce's 80,000 pages of unpublished manuscripts as well as the 12,000 already in print elsewhere. This first volume gives unique insight into the early development of Peirce's ideas on logic and philosophy of science. GHM

History, P, L. Charles Babbage: Pioneer of the Computer. Anthony Hyman. Princeton U Pr, 1982, xv + 287 pp, $25. [ISBN: 0-691-08303-7] This biography contains a remarkable and fascinating compilation of information, episodes and anecdotes from the life (1791-1871) of this leading (but unsuccessful) British advocate of the systematic use of science in industry. Babbage is now most famous as the unheeded inventor of the difference and analytical engines, precursors of today's computers. GHM

History, P, L. The Evolution of Dynamics: Vibration Theory from 1687 to 1742. John T. Cannon, Sigalia Dostrovsky. Stud. in History of Math. & Physical Sci., V. 6. Springer-Verlag, 1981, ix + 184 pp, $39. [ISBN: 0-387-90626-6] A detailed guide to the major works in vibration theory, from Newton to Euler and the Bernoullis, which gave rise to the subject of dynamics of systems with many degrees of freedom. GHM

History, P, L. The Prehistory of the Theory of Distributions. Jesper Lützen. Stud. in the History of Math. & Physical Sci., V. 7. Springer-Verlag, 1982, viii + 232 pp, $42. [ISBN: 0-387-90647-9] A concise, well-documented survey of several streams of mathematical thought that were eventually subsumed (in 1950) in Schwartz's theory of distributions: duality in functional analysis, generalized differentiation, Sobolev's functionals, the generalized Fourier transform, delta-functions, and de Rham's currents. LAS

History, S(18), P, L*. 100 Years of Mathematics: A Personal Viewpoint. George Temple. Springer-Verlag, 1981, xvi + 316 pp, $32. [ISBN: 0-387-91192-8] A synopsis of key elements in mathematical research, from the mid-nineteenth to the mid-twentieth century, touching on the work of over 300 mathematicians in as many pages. Offers an intellectual scaffolding for the algebra-analysis-geometry paradigm of graduate school mathematics, more akin to an encyclopedic dictionary than to a historical treatise. Contains over 600 references to original papers. LAS

Foundations, P. Regressive Sets and the Theory of Isols. Thomas G. McLaughlin. Lect. Notes in Pure & Appl. Math., V. 66. Dekker, 1982, vi + 371 pp, $49.50 (P). [ISBN: 0-8247-1337-0] An esoteric

blend of recursion theory, set theory and algebra applied to isols, a type of degree of isolated sets. Addressed to specialists only. The first unified presentation of the subject. GHM

Foundations, P, L. Ordered Sets. Ed: Ivan Rival. D Reidel Pub, 1982, xviii + 966 pp, $99. [ISBN: 90-277-1396-0] Proceedings of the 1981 NATO Symposium on Ordered Sets. Twenty-three survey articles touching upon all areas of current research on ordered sets, from set theory and recursion, to lattice theory and enumeration problems, to applications in computer and social sciences. Lists nearly 200 unsolved research problems. GHM

Foundations, P, L. Three Concepts of Time. Kenneth G. Denbigh. Springer-Verlag, 1981, viii + 180 pp, $19.80 (P). [ISBN: 0-387-10757-6] Philosophical analysis of time as a composite notion having three quite different sources: the time of consciousness having a definite "present;" the directionless, presentless time of physics; and the irreversible, presentless time of thermodynamics and evolutionary sciences. Seems well written and accessible to nonspecialists. GHM

Combinatorics, P. Combinatorial and Geometric Structures and their Applications. Ed: A. Barlotti. Math. Stud., V. 63. Elsevier North-Holland, 1982, viii + 292 pp, $44.25 (P). [ISBN: 0-444-86384-2] A collection of twenty-two papers on finite geometric structures. Some articles are expository; most are research reports. A high price for what amounts to a journal issue. SS

Discrete Mathematics, T(13-14: 1), L. Understanding Finite Mathematics. James Radlow. Prindle, Weber & Schmidt, 1982, xi + 624 pp. [ISBN: 0-87150-328-X] The author has "tried to be informal and unintimidating." He is, but the price is a sometimes painfully slow pace. Examples: matrices first appear on page 112; linear programming on page 212; conditional probability on page 392. Usual applications to Markov chains, game theory, finance. Exercises, answers, tables. JS

Number Theory, T(18), P. Introduction to Cyclotomic Fields. Lawrence C. Washington. Grad. Texts in Math., V. 83. Springer-Verlag, 1982, xi + 389 pp, $39. [ISBN: 0-387-90622-3] A beautiful, up-to-date survey of the subject; ideal for graduate students, experts in number theory, and everyone in between. Highlights include p-adic L-functions, Stickelberger's theorems, measures and distributions, and Iwasawa theory. SG

Number Theory, S*(15-17), P*, L*. Unsolved Problems in Number Theory. Richard K. Guy. Unsolved Prob. in Intuitive Math., V. 1. Springer-Verlag, 1981, xviii + 161 pp, $18. [ISBN: 0-387-90593-6] First of a series of volumes presenting conjectures and unsolved problems (with extensive references) intended for students, teachers and serious amateur math buffs. This volume covers primes, divisibility, sequences, Diophantine equations, and additive number theory. Problems are introduced with definitions, examples and full references. Presumes a good first course in number theory for basic definitions and theorems. LAS

Number Theory, P. Arithmetic on Modular Curves. Glenn Stevens. Progress in Math., V. 20. Birkhauser Boston, 1982, xvii + 214 pp, $15. [ISBN: 3-7643-3088-0] An exposition of the author's research on the connection between arithmetic properties of certain modular curves and special values of associated L-functions. SG

Linear Algebra, T(15-17: 1), L. Advanced Matrix Theory for Scientists and Engineers. Assem S. Deif. Halsted Pr, 1982, x + 241 pp, $34.95 (P). [ISBN: 0-470-27316-1] Covers standard topics in applied linear algebra, with an additional chapter on perturbation theory. Reasonable number of examples and many exercises, few of which are routine. Primarily intended for engineers and other practical users. JRG

Algebra, S(18), P. Products of Reflections in U(p,q). Dragomir Z. Djoković, Jerry G. Malzan. Memoirs No. 259. AMS, 1982, vi + 82 pp, $5.20 (P). This monograph deals with products of reflections in the unitary group U(p,q). For each $A \in U(p,q)$ with $\det(A) = \pm 1$ the authors have determined the minimum number of positive reflections needed to represent A. CEC

Algebra, T(14-15), L.** Introduction to Algebra. A.I. Kostrikin. Trans: Neal Koblitz. Springer-Verlag, 1982, xiii + 575 pp, $28 (P). [ISBN: 0-387-90711-4] A Soviet abstract algebra text. The first part contains six chapters: Sources of Algebra; Vector Spaces; Determinants; Algebraic Structures (groups, rings, fields); Complex Numbers and Polynomials; Roots of Polynomials. The second part delves more deeply into these topics with chapters on groups, representation theory and fields, rings and modules. Contains interesting applications, historical notes, and many perceptive comments. The organization is somewhat unusual, but the exposition is firm, concrete, and clear. This is an interesting algebra book, worth using as a text. SG

Algebra, P. Geometry of Coxeter Groups. Howard Hiller. Research Notes in Math., No. 54. Pitman Pub, 1982, 213 pp, $19.95 (P). [ISBN: 0-273-08517-4] These notes cover the geometry of flag manifolds from the algebraic and combinatorial point of view. The principal topics are the invariant theory of Coxeter groups and the combinatorics of Bruhat order, with the former constituting the main theme in studying a variety of topics that bring together geometry, group theory and combinatorics. SS

Algebra, T(18), S, P. Theory of Group Representations. M.A. Naimark, A.I. Stern. Trans: Elizabeth & Edwin Hewitt. Grund. der math. Wissenschaften, B. 246. Springer-Verlag, 1982, ix + 568 pp, $59. [ISBN: 0-387-90602-9] The first two chapters, with ample examples and exercises, are devoted to the algebraic aspects of the theory. Later chapters are concerned with topological groups, Lie groups

and Lie algebras, and their representations. Throughout, the theory is limited to finite-dimensional representations. LCL

Algebra, P. The Cohomology of Chevalley Groups of Exceptional Lie Type. Samuel N. Kleinerman. Memoirs No. 268. AMS, 1982, viii + 82 pp, \$5 (P).

Algebra, P. Finite Groups III. B. Huppert, N. Blackburn. Grund. der math. Wissenschaften, B. 243. Springer-Verlag, 1982, ix + 454 pp, \$59. [ISBN: 0-387-10633-2] The last of a three-volume series. Contains three long chapters on local finite group theory, Zassenhaus groups, and multiply transitive permutation groups. SG

Algebra, P. C*-bundles and Compact Transformation Groups. Bruce D. Evans. Memoirs No. 269. AMS, 1982, vii + 63 pp, \$4 (P). [ISBN: 0-8218-2269-1] Let G be a compact Lie group which acts smoothly on a C^∞ manifold X. The author develops a geometric theory for the structure of the associated C*-algebra, C*(G,X). LCL

Calculus, T(15-16: 1), S*, L. A Brief on Tensor Analysis. James G. Simmonds. Undergrad. Texts in Math. Springer-Verlag, 1982, xi + 92 pp, \$18.80. [ISBN: 0-387-90639-8] From two-dimensional vectors to Christoffel symbols, covariant derivatives and the divergence theorem--all in under 100 pages! A spare, informal text, with numerous exercises. Excellent supplement to advanced mathematics and physics courses; also suitable as a seminar text. LAS

Calculus, T(13). Calculus with Applications. Daniel I. Auvil. Addison-Wesley, 1982, xii + 558 pp, \$23.95. [ISBN: 0-201-10063-0] Should be called "Calculus for the Social Sciences" or "Calculus Without Trigonometry" (except for the last chapter). One of the less superficial books of this genre. AWR

Calculus, S(14-15), L. Mathematics 4 Checkbook. J.O. Bird, A.J.C. May. Butterworths, 1981, viii + 216 pp, \$22.95; \$9.95 (P). [ISBN: 0-408-00660-9; 0-408-00612-9] A collection of 300 worked problems and 500 further problems with answers, covering review of hyperbolic functions, calculus through elementary differential equations, Fourier series, Laplace transform, and statistics (correlation and regression). LCL

Real Analysis, S(18), P. Hardy Spaces on Homogeneous Groups. G.B. Folland, E.M. Stein. Math. Notes, V. 28. Princeton U Pr, 1982, xii + 284 pp, \$13 (P). The classical Hardy spaces, conceived in the context of complex analysis, developed as important tools of Fourier analysis. Modern H^p theory has moved further in the direction of real analysis, via harmonic Hardy spaces, the duality of H^1 and BMO, and certain decomposition theorems for H^p functions. This monograph develops real variable H^p theory on homogeneous groups. PZ

Real Analysis, T(17-18: 1, 2), S, P. The Theory of Generalised Functions. D.S. Jones. Cambridge U Pr, 1982, xiii + 539 pp, \$59.50. [ISBN: 0-521-23723-8] A thorough, readable introduction with exercises and worked examples to the Temple-Lighthill theory of distributions and their applications. "Generalized functions" are continuous operators of the Schwartz class; "weak functions" are distributions. Starts from advanced calculus--necessary convergence and Lebesgue integration theory is developed, though quickly, in the text. One-variable theory precedes that in R^n. PZ

Real Analysis, T*(15-16: 1, 2), S*, L. Introductory Problem Courses in Analysis and Topology. Edwin E. Moise. Universitext. Springer-Verlag, 1982, 94 pp, \$12 (P). [ISBN: 0-387-90701-7] Short, self-contained, yellow-brick road, to "baby" real-variable theory and elementary set-theoretic topology. Each chapter contains relevant definitions and theorems which the student is asked to prove. In the problem sets, the student is asked to resolve the status of the "propositions" which follow--providing proofs or counterexamples. This research-like approach should prove to be very attractive, enjoyable, and rewarding to everyone concerned. LCL

Complex Analysis, S(17-18), P. Factorization Theory of Meromorphic Functions and Related Topics. Ed: Chung-Chun Yang. Lect. Notes in Pure & Appl. Math., V. 78. Dekker, 1982, ix + 194 pp, \$27.50 (P). [ISBN: 0-8247-1834-8] Fourteen papers on questions in factorization of meromorphic functions on the plane. Factorization (in the sense of composition) for polynomials and entire functions is well understood by comparison with the meromorphic theory which, the editor asserts, is in its infancy. The final paper, by the editor, is a survey of progress in the area. PZ

Complex Analysis, T(18: 1, 2), S, P. Bounded Analytic Functions. John G. Garnett. Pure & Appl. Math. Academic Pr, 1981, xvi + 467 pp, \$59. [ISBN: 0-12-276150-2] Introduces and surveys results and, especially, techniques of bounded analytic functions on the unit disk. Assuming graduate complex and real analysis, the treatment is self-contained. Topics include BMO, interpolating sequences, and the corona construction. Much of the last five chapters is recent work, not elsewhere collected. With exercises, could be an intermediate graduate text. (Extended Review, August-September 1982.) PZ

Complex Analysis, P. Moduli of Families of Curves and Quadratic Differentials. G.V. Kuz'mina. Proc. of Steklov Inst. of Math., No. 139. AMS, 1982, vii + 231 pp, \$76 (P). [ISBN: 0-8218-3040-6] A thorough, self-contained exposition of the application of the method of extremal motions to geometric problems in function theory. Studies questions about minimal capacity and maximal diameter. LAS

Differential Equations, T(17-18: 1), S, P. Pseudo-Differential Operators. Hitoshi Kumano-go. MIT Pr, 1981, xviii + 455 pp, $60. [ISBN: 0-262-11080-6] Because of a wide range of applications, pseudo-differential operators have been shown to be a fruitful object of study. The approach is based on elementary calculus and Fourier transforms with the hope (probably vain) of becoming accessible to undergraduates. Three brisk chapters of theory lead to six chapters on significant applications. Formal in style, no exercises. Appendices, bibliographies, index. JS

Differential Equations, P. Evolution Equations and Their Applications. Ed: F. Kappel, W. Schappacher. Research Notes in Math., No. 68. Pitman Pub, 1982, xviii + 313 pp, $23.95 (P). [ISBN: 0-273-08567-0] Proceedings of the conference on differential equations and applications at the Volksbildungsheim Schlosz Retzhof, June 2-6, 1981. Topics covered include recent developments in the theory of semigroups and their applications to partial differential equations, integro-differential equations and delay equations. Applications to epidemics, population dynamics and control theory. Twenty papers in all. JK

Differential Equations, T(16-17: 1), P*. Hyperbolic Boundary Value Problems. Reiko Sakamoto. Transl: Katsumi Miyahara. Cambridge U Pr, 1982, ix + 210 pp, $34.50. [ISBN: 0-521-23568-5] On the existence and uniqueness of solutions of hyperbolic initial value problems in partial differential equations of higher orders. Prerequisites include some knowledge of complex variables and Lebesgue integration. Otherwise, self-contained. Carefully written and translated with several helpful translator footnotes. JK

Differential Equations, T(17), S, P. Solution of Differential Equations by Means of One-Parameter Groups. J.M. Hill. Research Notes in Math., No. 63. Pitman Pub, 1982, 161 pp, $18.95 (P). [ISBN: 0-85896-893-2] Advanced teaching material. Intended for applied mathematicians and engineers concerned with obtaining solutions of differential equations. Some new material on group approach to linear equations. Numerous examples, exercises and problems. JK

Differential Equations, P. Applications of Centre Manifold Theory. Jack Carr. Appl. Math. Sci., No. 35. Springer-Verlag, 1981, 142 pp, $14 (P). [ISBN: 0-387-90577-4] A centre manifold is a type of invariant manifold for a system of differential equations. Applications to bifurcation theory. JG

Differential Equations, P. Homology and Dynamical Systems. John M. Franks. CBMS Reg. Conf. Ser. in Math., No. 49. AMS, 1982, viii + 120 pp, $14 (P). [ISBN: 0-8218-1700-0] Notes of a 1980 conference. The basic theme is the connection between the dynamical configuration and the homological structure of a system. SG

Differential Equations, T(17-18), P, L. Partial Differential Equations, Fourth Edition. Fritz John. Appl. Math. Sci., V. 1. Springer-Verlag, 1982, x + 249 pp, $18. [ISBN: 0-387-90609-6] A few new proofs and examples; more problems, and corrections from previous editions. (Second Edition, TR, August-September 1975; Third Edition, TR, April 1979.) LCL

Differential Equations, P. A Method of Generalized Characteristics. Marc A. Berger, Alan D. Sloan. Memoirs No. 266. AMS, 1982, v + 37 pp, $4 (P). The classical and Brownian methods of characteristics are generalized to analyze evolution equations of arbitrary order. The method produces explicit solutions involving ordinary integration or recursion; the formal rules of manipulation are accessible to anyone with an advanced calculus background. LCL

Differential Equations, P. Singular Systems of Differential Equations II. S.L. Campbell. Research Notes in Math., No. 61. Pitman Pub, 1982, 234 pp, $19.95 (P). [ISBN: 0-273-08516-6] A continuation of the first volume (see TR, February 1982). In this book, more emphasis is placed on analytic and numerical procedures. LCL

Differential Equations, T(13-14), S. Modelling with Differential Equations. D.N. Burghes, M.S. Borrie. Math. & its Appl. Halsted Pr, 1981, 172 pp, $24.95 (P). [ISBN: 0-470-27101-9] A series of elementary (and generally well-known) applications of elementary differential equations: carbon dating, inhibited growth, art forgeries, pollution, oscillations, planetary motion, chemical kinetics, competing species, arms races, epidemics, and much more. Both the mathematics and the models are presented in simplified form, suitable for beginners. (TR, Hardcover, March 1982.) LAS

Differential Equations, P. Lecture Notes in Mathematics-942: Theory and Applications of Singular Perturbations. Ed: W. Eckhaus, E.M. de Jager. Springer-Verlag, 1982, v + 363 pp, $18.50 (P). [ISBN: 0-387-11584-6] Proceedings of an August 1981 Oberwolfach conference. LAS

Numerical Analysis, L. Methods of Numerical Mathematics, Second Edition. G.I. Marchuk. Trans: Arthur A. Brown. Appl. of Math., No. 2. Springer-Verlag, 1982, xiii + 510 pp, $69. [ISBN: 0-387-90614-2] This work is primarily a survey of modern numerical methods having widespread application. It is a significant revision of the First Edition (TR, January 1977), and incorporates new material throughout. AO

Numerical Analysis, P. Boundary Element Methods. Ed: C.A. Brebbia. Springer-Verlag, 1981, xxiv + 622 pp, $59. [ISBN: 0-387-10816-5] Proceedings of an international seminar held at Irvine, California during July 1981. LAS

Numerical Analysis, S(17-18), P. Lecture Notes in Mathematics-916: Zweidimensionale, interpolierende Lg-Splines und ihre Anwendungen. Karl-Ulrich Grusa. Springer-Verlag, 1982, vii + 238 pp, $11.10 (P). [ISBN: 0-387-11213-8] An account, for specialists, of the theory and some applications of two-dimensional Lg-Splines. JD-B

Numerical Analysis, S(17-18), P. Numerical Methods of Approximation Theory, V. 6. Ed: L. Collatz, G. Meinardus, H. Werner. ISNM 59. Birkhauser Boston, 1982, 265 pp, $29.95. [ISBN: 3-7643-1304-8] Papers presented at an Oberwolfach workshop, January 1981. JD-B

Numerical Analysis, S(17-18), P. Numerical Treatment of Free Boundary Value Problems. Ed: J. Albrecht, L. Collatz, K.-H. Hoffmann. ISNM 58. Birkhauser Boston, 1982, viii + 349 pp, $35. [ISBN: 3-7643-1277-7] Papers presented at an Oberwolfach workshop in November 1980. JD-B

Numerical Analysis, P. Proceedings of the 1982 Army Numerical Analysis and Computers Conference. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC), 1982, xiv + 606 pp, (P). Papers from a conference at Vicksburg, Mississippi, February 3-4, 1982. LAS

Functional Analysis, P. Lecture Notes in Mathematics-923: Functional Analysis in Markov Processes. Ed: M. Fukushima. Springer-Verlag, 1982, 307 pp, $18 (P). [ISBN: 0-387-11484-X] This volume contains 15 articles based on talks given during the summer of 1981 in Katata and Kyoto, Japan. AO

Functional Analysis, P. L^p boundedness of Fourier Integral Operators. R. Michael Beals. Memoirs No. 264. AMS, 1982, viii + 57 pp, $4 (P). Results concerning the LP boundedness of a class of Fourier integral operators of interest in the theory of hyperbolic partial differential equations. AO

Functional Analysis, P. Factorization of Matrix Functions and Singular Integral Operators. K. Clancey, I. Gohberg. Operator Theory, V. 3. Birkhauser Boston, 1981, x + 234 pp, $17.55. [ISBN: 3-7643-1297-1] Selection of material emphasizes connections between the factorization of non-singular matrix functions along closed contours and the solution of systems of singular integral equations. Generalized factorization and connections with the theory of singular integral operators are also discussed. AWR

Functional Analysis, S(18), P. Riesz and Fredholm Theory in Banach Algebras. B.A. Barnes, et al. Res. Notes in Math., V. 67. Pitman Pub, 1982, 123 pp, $16.95 (P). [ISBN: 0-273-08563-8] Extends the theory of Riesz and Fredholm linear operators on Banach spaces to the setting of Banach and C* algebras. Apart from an initial chapter on basic results in Banach spaces, most proofs are given in full. Many results are new. For the specialist. PZ

Functional Analysis, T*(16-18: 1, 2), S*, P. Theorems and Problems in Functional Analysis. A.A. Kirillov, A.D. Gvishiani. Trans: Harold H. McFaden. Problem Books in Math. Springer-Verlag, 1982, ix + 347 pp, $38. [ISBN: 0-387-90638-X] The theory covered (in only 135 pages) is about that of a typical one-year graduate functional analysis course with measure theory. Over 800 conveniently organized problems follow, ranging from entirely straightforward to very difficult. 23 "essential" exercises are identified in each subsection. Part III contains hints for every exercise. Inviting and invaluable graduate text or companion text. PZ

Functional Analysis, S(18), P. Lecture Notes in Mathematics-934: Quadrature Domains. Makoto Sakai. Springer-Verlag, 1982, 133 pp, $8.80 (P). [ISBN: 0-387-11562-5] Discussion of existence and uniqueness for positive measures on complex quadrature domains for various (analytic, harmonic, subharmonic) classes of functions. Applications to Dirichlet integrals, Gaussian curvature, Helle-Shaw flows. Bibliography, index. JS

Functional Analysis, T(17-18: 2), S, P. Functional Analysis, Second Edition. L.V. Kantorovich, G.P. Akilov. Transl: Howard L. Silcock. Pergamon Pr, 1982, xiv + 589 pp, $25 (P); $100. [ISBN: 0-08-026486-7; 0-08-023036-9] Second edition of the 1959 original. An encyclopedic treatment of the basics of functional analysis, and, notably, certain applications to real analysis--integral operators, functional equations, the fixed-point principle, etc. Level appropriate to a graduate text, though no exercises. Certainly a valuable reference work. PZ

Functional Analysis, T*(16-17: 1, 2), S*, L. Linear Operator Theory in Engineering and Science. Arch W. Naylor, George R. Sell. Appl. Math. Sci., V. 40. Springer-Verlag, 1982, xv + 624 pp, $28. [ISBN: 0-387-90748-3] Functional analysis in form suitable for engineers, scientists and applied mathematicians. In definition-theorem-proof format, but very readable with much attention to motivation. Numerous illustrative examples. Bountiful supply of problems which extend the text. Appendices on integration and measure theory and on probability spaces and stochastic processes. Flexible to accommmodate different audiences. Originally published in 1971; a welcome new edition. JK

Analysis, T(16-17), S, P, L. Calculus of Variations and Partial Differential Equations of the First Order, Second Edition. C. Carathéodory. Transl: Robert B. Dean. Chelsea Pub, 1982, xix + 402 pp, $25. A new edition of a great classic, now combined into one volume with known misprints corrected. (First Edition, Volume I, TR, April 1967; Volume II, TR, April 1968.) AWR

Analysis, T(18: 1), S, P. Hankel Operators on Hilbert Space. S.C. Power. Research Notes in Math., No. 64. Pitman Pub, 1982, 87 pp, $13.95 (P). [ISBN: 0-273-08518-2] An account of modern aspects of

the operator and spectral theory of Hankel operators, with particular attention to those arising from recent progress in function theory on the unit circle. Includes detailed proofs of earlier results and recent research. Problems and an extensive bibliography. CEC

Analysis, P. Ergodic Theory and Dynamical Systems II: Proceedings Special Year, Maryland 1979-80. Ed: A. Katok. Progress in Math., V. 21. Birkhauser Boston, 1982, xi + 210 pp, $15. [ISBN: 3-7643-3096-1] Seven survey lectures, complementing Part I (Progress in Math., V. 10; TR, May 1982), concluding the proceedings of the 1979-80 special year at Maryland. LAS

Algebraic Geometry, P. Automorphic Forms, Representation Theory and Arithmetic. S. Gelbart, et al. Tata Inst. of Fund. Res. Stud. in Math., No. 10. Springer-Verlag, 1981, 355 pp, $17.40 (P). [ISBN: 0-387-10697-9] Proceedings of the international colloquium held at the Tata Institute of Fundamental Research, Bombay, India, January 8-15, 1979. JAS

Algebraic Geometry, T(16-17: 2), S. Ebene Algebraische Kurven. E. Brieskorn, H. Knörrer. Birkhauser Boston, 1981, xi + 964 pp, $24.95. [ISBN: 3-7643-3030-9] A rich and leisurely introduction to the theory of algebraic curves in the complex projective plane. Many sketches, historical references, specific examples. Bibliography, but no exercises. JD-B

Algebraic Geometry, P. Lecture Notes in Mathematics-930: Théorie de Dieudonné, Cristalline II. P. Berthelot, L. Breen, W. Messing. Springer-Verlag, 1982, xi + 261 pp, $14.90 (P). [ISBN: 0-387-11556-0] Generalizes "Dieudonné theory" to the case of p-divisible groups and arbitrary schemes of positive characteristic. SG

Algebraic Geometry, T(18), P. Lecture Notes in Mathematics-935: Simple Morphisms in Algebraic Geometry. Richard Sot. Springer-Verlag, 1982, iv + 146 pp, $8.50 (P). [ISBN: 0-387-11564-1] An introductory text which assumes only basic topology, algebra, and commutative algebra. Describes various criteria for an algebra to be simple over its base field. Concludes with a discussion of simple morphisms of preschemes. A clearly written text. SG

Algebraic Geometry, P. Lecture Notes in Mathematics-917: Brauer Groups in Ring Theory and Algebraic Geometry. Ed: F. van Oystaeyen, A. Verschoren. Springer-Verlag, 1982, viii + 300 pp, $18 (P). [ISBN: 0-387-11216-2] The proceedings of a conference held at the University of Antwerp in August 1981. LCL

Algebraic Geometry, P. Lecture Notes in Mathematics-946: Classes Unipotentes et Sous-groupes de Borel. Nicolas Spaltenstein. Springer-Verlag, 1982, ix + 259 pp, $14 (P). [ISBN: 0-387-11585-4] A study of the fixed points of the action of an affine algebraic group on the variety of its Borel subgroups. SG

Differential Geometry, P. Global Differential Geometry of Surfaces. A. Svec. Kluwer Boston, 1981, 153 pp, $28.50. [ISBN: 90-277-1295-6] An exposition of the author's research on global properties of surfaces; the principal results describe sufficient conditions for a surface to be a sphere or a plane. SG

Differential Geometry, T(18), S, P. Cusps of Gauss Mappings. Thomas Banchoff, Terence Gaffney, Clint McCrory. Pitman Pub, 1982, 88 pp, $12.95 (P). [ISBN: 0-273-08536-0] A self-contained, readable presentation of the authors' research on geometric characterizations of cusps of Gauss mappings of surfaces in R^3. Contains many pictures, examples, and references. SG

Geometry, P. The Geometric Vein: The Coxeter Festschrift. Ed: Chandler Davis, Branko Grünbaum, F.A. Sherk. Springer-Verlag, 1981, viii + 598 pp, $48. [ISBN: 0-387-90587-1] A listing of Coxeter's publications followed by a collection of papers from the May 1979 Coxeter symposium (held at the University of Toronto). JNC

Geometry, T(16-17: 1), S, P. Symmetrien von Ornamenten und Kristallen. Michael Klemm. Springer-Verlag, 1982, vii + 214 pp, $16 (P). [ISBN: 0-387-11644-3] A thorough introduction to mathematical crystallography via the summetry groups of ornaments and crystals. LAS

Algebraic Topology, T(17-18: 1, 2). Differential Forms in Algebraic Topology. Raoul Bott, Loring W. Tu. Grad. Texts in Math., No. 82. Springer-Verlag, 1982, xiv + 331 pp, $29.80. [ISBN: 0-387-90613-4] A serious introduction to algebraic topology using differential forms and the De Rham theory as the principal tool. The implied limitation to spaces which are real manifolds does not appear to limit seriously the breadth of material covered. Good index and exercises. JAS

Algebraic Topology, P. Induction Theorems for Groups of Homotopy Manifold Structures. Andrew J. Nicas. Memoirs No. 267. AMS, 1982, vi + 108 pp, $6 (P). A study of group structure in and extension of the surgery exact sequence of compact oriented manifolds with boundary. JG

Algebraic Topology, T(17-18). Éléments D'Analyse, Tome IX. J. Dieudonné. Gauthier-Villars, 1982, xviii + 380 pp, 320 FF (P). [ISBN: 2-04-011499-8] An introduction to algebraic and differential topology. Special topics include Chern and Stiefel-Whitney classes, Hodge theory, the Atiyah-Bott-Lefschetz formula, cohomology of Lie groups and Grassmannians. Many exercises. SG

Topology, P. Chain Conditions in Topology. W.W. Comfort, S. Negrepontis. Tracts in Math., No. 79. Cambridge U Pr, 1982, xiii + 300 pp, $39.50. [ISBN: 0-521-23487-5] A collection of new results by

the authors concerning various considerations related to the cardinality of the open sets of a topological space. The work involves recent modifications of combinatorial tools developed in the 1960's by Erdős and Rado. JAS

Topology, P. Embedding Coverings Into Bundles with Applications. P.F. Duvall, L.S. Husch. Memoirs No. 263. AMS, 1982, iv + 53 pp, $4 (P). A solution to the problem of when a finite regular covering of a closed n-manifold is homotopic to an embedding in an n-plane bundle over the manifold. This solution is then applied to a study of embedding n-manifold-like continua (up to shape) into 2n-space. JAS

Topology, S(17), P, L. Geometric Theory of Dynamical Systems, An Introduction. Jacob Palis, Jr., Welington de Melo. Transl: A.K. Manning. Springer-Verlag, 1982, xii + 198 pp, $28. [ISBN: 0-387-90668-1] An introduction to the structural stability and genericity of vector fields. Assumes background in differential equations, differentiable manifolds, linear algebra, and Banach spaces. JG

Topology, P, L. Quasi-Uniform Spaces. Peter Fletcher, William F. Lindgren. Pure & Appl. Math., V. 77. Dekker, 1982, viii + 216 pp, $29.75 (P). [ISBN: 0-8247-1839-9] Compilation and exposition of the widely scattered literature on quasi-uniformities and quasi-proximities, intended to encourage the use of those structures in general topology. Includes historical notes, lengthy bibliography, index, and 22 unsolved problems. In short, a valuable resource. GHM

Operations Research, T(16-18: 1), L. Multiobjective Decision Analysis with Engineering and Business Applications. Ambrose Goicoechea, Don R. Hansen, Lucien Duckstein. Wiley, 1982, xvii + 519 pp, $34.95. [ISBN: 0-471-06401-7] An introduction to multi-objective optimization problems. Considers both continuous and discrete problems and presents a number of algorithms for each. Provides Fortran computer programs for two of the methods in an appendix. AO

Operations Research, S(15-16), L. Network Optimisation Practice: A Computational Guide. David K. Smith. Halsted Pr, 1982, 237 pp, $59.95. [ISBN: 0-470-27347-X] An elementary introduction to several standard network optimization problems and algorithms for their solution. Listings of Basic and Pascal implementations of many of these algorithms are included. AO

Optimization, P. Modern Applied Mathematics: Optimization and Operations Research. Ed: Bernhard Korte. North-Holland, 1982, ix + 693 pp, $130.25. [ISBN: 0-444-86134-3] A collection of 17 articles based on talks presented during a 1979 summmer school organized by the University of Bonn. The articles provide state-of-the-art surveys of a number of areas of optimization theory and operations research. AO

Optimization, T*(16-17: 1, 2), L*. Optimization Techniques: An Introduction. L.R. Foulds. Undergrad. Texts in Math. Springer-Verlag, 1981, xi + 502 pp, $36. [ISBN: 0-387-90586-3] This textbook provides an introduction to some of the most important optimization techniques at an undergraduate level. The topics covered include linear programmming, integer programmming, network analysis, dynamic programming, classical optimization, and nonlinear programming. AO

Optimization, P. Mathematical Techniques of Optimization, Control and Decision. Ed: J.P. Aubin, A. Bensoussan, I. Ekeland. Birkhauser Boston, 1981, viii + 212 pp, $24.95. [ISBN: 3-7643-3032-5] Report from the Research Center in the Mathematics of Decision Making. Collection of essays surveying methods and applications in systems theory and decision making. JRG

Optimization, T(16-17: 2), P. Control, Identification, and Input Optimization. Robert Kalaba, Karl Spingarn. Math. Concepts & Methods in Sci. & Eng., V. 25. Plenum Pr, 1982, xi + 431 pp, $39.50. [ISBN: 0-306-40847-3] Major divisions of the book are: I. Optimal control and methods for numerical solutions; II. System identification; III. Optimal inputs for system identification. This attractive book is written to be accessible to students with no previous knowledge of optimal control theory or numerical methods, but it will be a useful reference to those in control engineering. AWR

Optimization, T(17: 2), P. Optimization Over Time: Dynamic Programming and Stochastic Control, V. 1. Peter Whittle. Wiley, 1982, xi + 317 pp, $46.95. [ISBN: 0-471-10120-6] The goal is to optimize decisions to be made in running a dynamic system, a subject variously called dynamic programming or stochastic control. Of interest to people in operations research and control engineering. AWR

Probability, P. Lecture Notes in Mathematics-920: Séminaire de Probabilités XVI, 1980/81. Ed: J. Azéma, M. Yor. Springer-Verlag, 1982, v + 622 pp, $31.10 (P). [ISBN: 0-387-11485-8] This volume, together with the supplement (Lecture Notes 921) on stochastic differential geometry, form the proceedings of the 1980-1981 seminar at the Université de Strasbourg. JAS

Probability, S(15-17), L. The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics. Ed: Kh. O. Ondar. Transl: Charles and Margaret Stein. Springer-Verlag, 1981, xvii + 181 pp, $32. [ISBN: 0-387-90585-5] Translation of a 1977 Russian book containing much of the correspondence between these two from 1910-1917, together with editorial comments. Provides an interesting insight into Markov's personality, as well as valuable historical material. RSK

Probability, P. Lecture Notes in Mathematics-929: Ecole d'Eté de Probabilités de Saint-Flour X-1980. J.M. Bismut, L. Gross, K. Krickeberg. Springer-Verlag, 1982, x + 313 pp, $17.40 (P). [ISBN: 0-387-11547-1] Three courses from the 1980 Saint-Flour summer school in probability theory:

mechanics and stochastic optimization, equilibrium thermodynamics, and punctual processes. LAS

Statistics, S(16-17), P. Computing in Statistical Science through APL. Francis John Anscombe. Ser. in Stat. Springer-Verlag, 1981, xv + 426 pp, $24.80. [ISBN: 0-387-90549-9] The stated purpose of this book is to interest statisticians in the use of the programming language APL. The first third of the book presents a description of APL while the remainder presents several examples of statistical analysis using APL. AO

Statistics, T(16-18: 1, 2), S, P, L. The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory. Z. Govindarajulu. Math. & Management Sci. Ser., V. 5. American Sci Pr, 1981, ix + 680 pp, $48.50 (P). [ISBN: 0-935950-02-8] A comprehensive treatment of sequential testing and estimation. Presupposes courses in probability, statistics, and advanced calculus. FLW

Statistics, T(16-18: 1, 2), S, P, L. Concepts of Nonparametric Theory. John W. Pratt, Jean D. Gibbons. Ser. in Stat. Springer-Verlag, 1981, xvi + 462 pp, $38. [ISBN: 0-387-90582-0] Could serve as a text for a first or second course on nonparametric tests. Most of the material presupposes only a calculus-based probability and statistics course. Includes "considerable philosophical and methodological discussion." FLW

Statistics, S(15-17), P*, L. Circular Statistics in Biology. Edward Batschelet. Math. in Biology. Academic Pr, 1981, xvi + 371 pp, $69.50. [ISBN: 0-12-081050-6] Deals with analysis of data points distributed on a circle (e.g., directions in a plane, or in some cases other cyclical phenomena). Part I provides a thorough coverage of the descriptive and inferential tools of this often neglected area of statistics, while Part II contains the mathematics required. Good set of references. RSK

Statistics, S*(15-17), L. The Making of Statisticians. Ed: J. Gani. Springer-Verlag, 1982, viii + 263 pp, $19.80. [ISBN: 0-387-90684-3] Interesting collection of 16 autobiographical essays by pioneers (born 1897-1917) in the area of modern statistics. Includes sketches of their lives, explanations of how they became interested in probability and statistics, accounts of their major contributions, and some predictions about the future. RSK

Statistics, P. Lecture Notes in Statistics-14: GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Ed: Robert Gilchrist. Springer-Verlag, 1982, 188 pp, $12.50 (P). [ISBN: 0-387-90777-7]

Statistics, T(15-16: 1, 2), S. Introduction to Probability Theory and Statistical Inference, Third Edition. Harold J. Larson. Wiley, 1982, xi + 637 pp, $27.95. [ISBN: 0-471-05909-9] Presupposes one year of calculus. This edition has new material on descriptive statistics, the bivariate normal distribution, the F-distribution, and nonparametric procedures. (TR, First Edition, October 1969; Extended Review, November 1972; TR, Second Edition, December 1974.) FLW

Statistics, S(16-18), P. Lecture Notes in Statistics-9: Statistical Properties of the Generalized Inverse Gaussian Distribution. Bent Jørgensen. Springer-Verlag, 1982, 188 pp, $12 (P). [ISBN: 0-387-90665-7]

Statistics, T(13: 1), S, P, L. The Randomized Clinical Trial and Therapeutic Decisions. Ed: Niels Tygstrup, John M. Lachin, Erik Juhl. Statistics, V. 43. Dekker, 1982, xviii + 296 pp, $39.75. [ISBN: 0-8247-1856-9] Could serve as a text for doctors or potential doctors. Contributions by several M.D.'s and biostatisticians to a general treatment of randomized clinical trials. No exercise sets. FLW

Statistics, S(16-18), P. Connectivity in Multi-factor Designs: A Combinatorial Approach. Lothar Butz. Res. & Educ. in Math., V. 3. Heldermann Verlag, 1982, vii + 190 pp, 32 DM (P). [ISBN: 3-88538-203-2] Uses graph theoretic techniques to deal with experimental design problems. FLW

Statistics, T(13: 1), S, L. Statistics Without Tears: A Primer for Non-mathematicians. Derek Rowntree. Charles Scribner's, 1981, 199 pp, $6.95 (P). [ISBN: 0-684-17502-9] Introduces basic statistical concepts while expecting very little in the way of calculations to be done by the reader. No exercise sets, but questions for the reader occur in the body of the text. FLW

Statistics, S(16-18), P. Lecture Notes in Statistics-11: Random Coefficient Autoregressive Models: An Introduction. Des F. Nicholls, Barry G. Quinn. Springer-Verlag, 1982, v + 154 pp, $12 (P). [ISBN: 0-387-90766-1] Autoregressive linear models are those whose coefficients vary with time, adjusting automatically to changing conditions. This monograph develops the "quite tractable" theory of autoregressive models with random coefficients. LAS

Statistics, T(16-18: 1, 2), S, P. Estimation of Dependences Based on Empirical Data. Vladimir Vapnik. Transl: Samuel Kotz. Ser. in Stat. Springer-Verlag, 1982, xvi + 399 pp, $56. [ISBN: 0-387-90733-5] A study of the results of adopting the principle: "select, from an admissible set of functions, a function which fulfills a definite relationship between a quantity characterizing the quality of the approximation and a quantity characterizing the 'complexity' of the approximating function." FLW

Statistics, S*(14-18), P, L*. Encyclopedia of Statistical Sciences, Volume 2: Classification to Eye Estimate. Ed: Samuel Kotz. Wiley, 1982, viii + 613 pp, $75. [ISBN: 0-471-05547-6] Second volume of

a planned eight-volume set. (See TR, October 1982, of Series and Volume 1.) RSK

Computer Programming, T(13: 1). Problem Solving and Structured Programming in WATFIV. Frank L. Friedman, Elliot B. Koffman. Addison-Wesley, 1982, xvi + 539 pp, 16.95 (P). [ISBN: 0-201-10482-2] WATFIV is a dialect of Fortran. This textbook is an introduction to programming using the WATFIV language. It is based on the second edition of Problem Solving and Structured Programming in Fortran by the same authors (TR, December 1981). AO

Computer Programming, S(15-17), P, L. ADA, An Introduction: Ada Reference Manual (July 1980). Henry Ledgard. Springer-Verlag, 1981, vii + 373 pp, $16.80 (P). [ISBN: 0-387-90568-5] This volume consists of two parts. The first part is an introduction to the programming language Ada written for readers with experience in high-level language programming. Because there are no exercises, it is not suited for use as a textbook. The second part is a reproduction of the November 1980 printing of the Ada Reference Manual which defines the language. AO

Computer Programming, S(16-17), P. Programming in Ada. J.G.P. Barnes. Addison-Wesley, 1982, x + 340 pp, $17.95 (P). [ISBN: 0-201-13792-5] This textbook on the programming language Ada was written by one of the members of the Ada Design Team. It provides an introduction to nearly all aspects of the language and even incorporates some of the very recent changes to the language definition. AO

Computer Programming, T(13: 1). BASIC for Students: With Applications. Michael Trombetta. Addison-Wesley, 1981, xi + 291 pp, $9.95 (P). [ISBN: 0-201-07611-X] Suitable for a semester course. Clearly stated objectives. Uses flow charts. Subroutines under-emphasized. RBK

Computer Programming, T(13), S. Some Common Pascal Programs. Ed: Lon Poole. Osborne/McGraw-Hill, 1982, ix + 235 pp, $14.99 (P). [ISBN: 0-931988-73-X] Lists 76 complete Pascal programs. For each one the book provides a short introduction, the program listing and sample output. Programs are of a difficulty appropriate to a first course in programming. MS

Computer Programming, S(16), P. Synthetische Programmierung auf dem HP-41C/CV. W.C. Wickes. Heldermann Verlag, 1982, x + 137 pp, (P). [ISBN: 3-88538-800-6] German translation of the fifth edition of Synthetic Programming on the HP-41C CV. JD-B

Computer Programming, S(13), P. Problems for Computer Solution, Student Edition. Stephen J. Rogowski. Creative Computing Pr, 1979, v + 104 pp, $4.95 (P). [ISBN: 0-916688-13-5]; Problems for Computer Solution, Teacher Edition. 1980, iv + 181 pp, $9.95 (P). [ISBN: 0-916688-14-3] Ninety-plus problems arranged by mathematical topic, from arithmetic through calculus. Teacher's edition reproduces student page and includes sample printout, Basic listing and analysis. Good source although a few problems are more easily done without computer; Basic examples do not encourage structured approach. MW

Data Structures, T*(15-16: 1), L. Data Structures and Algorithms. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman. Addison-Wesley, 1983, xi + 427 pp, $26.95. [ISBN: 0-201-00023-7] This text is based on the material in the first six chapters of the well-known textbook The Design and Analysis of Computer Algorithms (TR, November 1975) by the same authors plus some additional material. It is designed to be used in a first course on data structures and algorithms. AO

Software Systems, S(16-17), P, L. Engineering a Compiler: VAX-11 Code Generation and Optimization. Patricia Anklam, et al. Digital Equip, 1982, xv + 269 pp, $24. [ISBN: 0-932376-19-3] This work relates the experiences of the authors while producing a commercial PL/I compiler. It emphasizes the software engineering aspects of compiler construction, particularly that of the code generation phase. AO

Computer Science, T(16-18: 1, 2), S, P, L. Fundamentals of Interactive Computer Graphics. James D. Foley, Andries van Dam. Addison-Wesley, 1982, xx + 664 pp, $34.95. [ISBN: 0-201-14468-9] A major four-part treatment covering both hardware and software: basics; mathematics, data structures, and display architecture; modern raster technology; making realistic synthetic photographs. Assumes only introductory material in programming, computer architecture, and data structures. JAS

Computer Science, T?(13-14: 2, 3), P?. Mathematics for Computing. G.P. McKeown, V.J. Rayward-Smith. Halsted Pr, 1982, xi + 428 pp, $24.95 (P). [ISBN: 0-470-27268-6] Potpourri of topics in discrete and continuous mathematics (including calculus, probability and abstract algebra) with emphasis on algorithms for computer implementation. Coverage seems too succinct for textbook use, too unreliable for reference (e.g., "Theorem 3.2.3. A necessary and sufficient condition for convergence of an alternating series $\{a_1-a_2+a_3-...\}$ is that $\{a_n\}$ is monotonically decreasing." There a "proof" is even sketched, though both implications are false). GHM

Computer Science, P. Lecture Notes in Computer Science-140: Automata, Languages and Programming. Ed: M. Nielsen, E.M. Schmidt. Springer-Verlag, 1982, vii + 614 pp, $27.60 (P). [ISBN: 0-387-11576-5] The proceedings of the Ninth International Colloquium on automata, languages and programming held at Aarhus University, from July 12-16, 1982. CEC

Computer Science, T(14: 1), L. Computer Architecture and Organization with Examples Using the PDP-11. Theodore H. Meyer, Jr. Dilithium Pr, 1982, x + 333 pp, $16.95 (P). [ISBN: 0-918398-55-X] Reviews the principles of computer organization using the PDP-11 as its model. Covers information representation, and the major components of a computer system--memory, I/O processor, bus structure.

About one-third of the book is devoted to MACRO-11 assembly language programming. MS

Computer Science, T*(17-18: 2), P. Principles of Database Systems, Second Edition. Jeffrey D. Ullman. Computer Sci Pr, 1982, vii + 484 pp, $24.95. [ISBN: 0-914894-36-6] Covers the complete field of the design and analysis of data base systems. Reviews the three major models--network, hierarchical, and relational--but concentrates most heavily on the relational model. Has an extensive mathematical treatment of many aspects of data base design including query optimization, distributed data base systems and design theory. Advanced book for the serious student. (TR, First Edition, August-September 1980.) MS

Control Theory, T*(17: 2), S, P*. Interconnected Dynamical Systems. Raymond A. DeCarlo, Richard Saeks. Electrical Engin. & Electronics, No. 10. Dekker, 1981, vi + 512 pp, $57.50. [ISBN: 0-8247-6639-3] Beginning graduate level textbook. Provides analytical tools for analysis and design of interconnected networks and systems. Discusses computational problems encountered in implementing theory. Well-illustrated. Problems. No answers. Numerous references plus extensive bibliography. JK

Systems Theory, P. Lecture Notes in Control and Information Sciences-38: System Modeling and Optimization. Ed: R.F. Drenick, F. Kozin. Springer-Verlag, 1982, xi + 893 pp, $52.40 (P). [ISBN: 0-387-11691-5] Proceedings of the 10th IFIP (International Federation of Information Processing) Conference, New York City, September 1981: five plenary addresses, five topical addresses, and scores of contributed papers on a very wide variety of modelling situations. LAS

Applications, S(15-16), L*. The Application of Mathematics in Industry. Ed: Robert S. Anderssen, Frank R. de Hoog. Kluwer Boston, 1982, xiv + 202 pp, $32.50. [ISBN: 90-247-2590-9] The proceedings of a one-day seminar held at the Australian National University in 1980, presenting a wide variety of industrial applications of mathematics. AO

Applications (Artificial Intelligence), T*(16-17: 1), S*, P. Progress in Pattern Recognition, Volume 1. Ed: Laveen N. Kanal, Azriel Rosenfeld. Elsevier North-Holland, 1981, vii + 391 pp, $53.25. [ISBN: 0-444-86325-7] A collection of 10 papers written especially for this proposed two-volume series in pattern recognition. The papers are survey or tutorial in nature describing recent progress in the area of data representation, algorithms for scene processing, and imaging software. For the most part the authors are all very well known people in the field and the book summarizes well the state of the art of pattern recognition as of 1980-81. MS

Applications (Biology), S(16-18), P. Lecture Notes in Biomathematics-45: Competition and Cooperation in Neural Nets. Ed: S. Amari, M.A. Arbib. Springer-Verlag, 1982, xiv + 441 pp, $26 (P). [ISBN: 0-387-11574-9] The last decade has seen an explosion of experimental research on the brain, yet the theory remains highly fragmented and uneven. These expository reports constitute the proceedings of a major conference to catalyze better integration of theory and experiment. Participants included brain surgeons, neurophysiologists, mathematicians, computer scientists, physicists. One cannot help but be impressed with the richness, the excitement, and the importance of this area of mathematical modeling. LCL

Applications (Biology), P. Lecture Notes in Biomathematics-44: Recognition of Pattern and Form. Ed: Duane G. Albrecht. Springer-Verlag, 1982, 225 pp, $15 (P). [ISBN: 0-387-11206-5] Proceedings of a conference held during March 1979 at the University of Texas, Austin. LAS

Applications (Cryptography), P, L*. Secure Communications and Asymmetric Cryptosystems. Ed: Gustavus J. Simmons. AAAS Selected Symposia Ser., No. 69. Westview Pr, 1982, x + 338 pp, $30. [ISBN: 0-86531-338-5] Papers presented at the 1980 AAAS annual meeting surveying the then-current state of "public key" cryptosystems that depend on an asymmetry determined by a computationally hard (NP-complete) problem. Includes papers on policy and applications, as well as details of various coding systems. LAS

Applications (Economics), P. Dynamic Analysis of Open Economies. Masanao Aoki. Academic Pr, 1981, xxvi + 341 pp, $39.50. [ISBN: 0-12-058940-0] An analysis of how economies with different structures are affected by common disturbances (e.g., inflation, exchange rates, internal and external balances, etc.), and how these disturbances spread throughout the world. The emphasis on dynamic analysis (employing variational and perturbation techniques) distinguishes this book from others with similar objectives. LCL

Applications (Engineering), T(16: 1), S, P*. Approximate Methods in Engineering Design. T.T. Furman. Math. in Sci. & Eng., V. 155. Academic Pr, 1981, viii + 388 pp, $59. [ISBN: 0-12-269960-2] Techniques and methods, along with case studies from mechanical engineering, in support of the philosophy "...approximate methods of calculations with a reasonable knowledge of the degree of accuracy involved can, in general, be much more valuable than time-consuming precise and rigorous mathematical analyses." Appropriate error and statistical analyses. JK

Applications (Engineering), P. Dynamics of Manipulation Robots: Theory and Application. M. Vukobratović, V. Potkonjak. Scientific Fund. of Robotics, No. 1. Springer-Verlag, 1982, xiii + 303 pp, $37.50. [ISBN: 0-387-11628-1] Computer methods for forming mathematical models of the dynamics of robots, with applications. Requires a knowledge of linear and nonlinear systems theory and a strong background in mechanism dynamics. LCL

Applications (Engineering), P. Control of Manipulation Robots: Theory and Application. M. Vukobratović, D. Stokić. Scientific Fund. of Robotics, No. 2. Springer-Verlag, 1982, xiii + 363 pp, $45. [ISBN: 0-387-11629-X] A continuation of the first volume in this series (see preceding review); this volume focuses on the control of robots. A two-level, suboptimal, decentralized procedure is described for implementing global control. Several illustrations to six-degree-of-freedom robots. LCL

Applications (Linguistics), P. A Grammar of English on Mathematical Principles. Zellig Harris. Wiley, 1982, xvi + 429 pp, $43.50. [ISBN: 0-471-02958-0] A new theory of English grammar based on a partial order between operator and argument words. "The partial order relation defines a subset of English sentences, which are not always colloquially customary, but in which all information possible to convey in the language can be expressed." LAS

Applications (Physics), P. Functional Integration: Theory and Applications. Ed: Jean-Pierre Antoine, Enrique Tirapegui. Plenum Pr, 1980, x + 355 pp, $42.50. [ISBN: 0-306-40573-3] The proceedings of an international workshop held in Louvain-la-Neuve in 1979. The papers present attempts to provide a rigorous mathematical foundation for the Feynman path integral as well as a number of applications in physics. AO

Applications (Physics), P. Lecture Notes in Mathematics-905: Differential Geometric Methods in Mathematical Physics. Ed: H.-D. Doebner, S.I. Andersson, H.R. Petry. Springer-Verlag, 1982, vi + 309 pp, $18 (P). [ISBN: 0-387-11197-2] Proceedings of an international conference held at the Technical University of Clausthal, West Germany, July 23-25, 1980. JAS

Applications (Physics), P. Stochasticity and Partial Order: Doubly Stochastic Maps and Unitary Mixing. Peter M. Alberti, Armin Uhlmann. Math. & Its Appl., V. 9. D Reidel Pub, 1982, 123 pp, $28.50. [ISBN: 90-277-1350-2] The classical theory of majorization (an important partial order relation on n-tuples of positive numbers with the same sum) is generalized to state spaces of W*-algebras, with applications to linear system theory, thermodynamics and stochastics. LCL

Applications (Physics), P. Point Group Symmetry Applications: Methods and Tables. Philip H. Butler. Plenum Pr, 1981, ix + 567 pp, $55. [ISBN: 0-306-40523-7] A useful reference for graduate studies in molecular chemistry and quantum mechanics. Starting with the assumption that the reader knows little or no group theory but is familiar with elementary quantum mechanics, the author proceeds to show how to understand and make best use of the tables which follow (which include all 3jm and 6j symbols for all point group schemes). LCL

Applications (Physics), S(18), P. Inverse Scattering Papers: 1955-1962. Irvin Kay, Harry W. Moses. Lie Groups: History, Frontiers & Appl., V. XII. Math Sci Pr, 1982, 305 pp, $30. [ISBN: 0-915692-32-5] A collection of thirteen related papers on inverse scattering and spectral theory from ´55 to ´63. Emphasizes heuristic approach and development of formalism as a tool for constructing explicit potentials in eigenvalue problems. JS

Applications (Physics), P. Les Incertitudes D´Heisenberg et L´Interprétation Probabiliste de la Mécanique Ondulatoire. Louis de Broglie. Gauthier-Villars, 1982, xlii + 304 pp, 220 FF. [ISBN: 2-04-015411-6] An account of de Broglie´s ideas on the probabilistic interpretation of wave mechanics. The main text, written between 1950 and 1952 but never before published, develops the generally-accepted theory (of Bohr, et al.). Notes written over the next 25 years show how his views shifted in the direction of causality and hidden variables. JD-B

Applications (Physics), T(15-17: 2), S, L. Cosmology, Physics, and Philosophy. Benjamin Gal-Or. Springer-Verlag, 1981, xv + 522 pp, $28. [ISBN: 0-387-90581-2] An extraordinary book, linking separate parts of cosmology and physics with modern results to provide the basis for a contemporary version of natural philosophy. Mixes personal philosophy and religious/mystical images with traditional science in a highly idiosyncratic manner. The author suggests this book as a basis for a core curriculum course, but much of it requires vector calculus for reasonable understanding. LAS

Applications (Physics), P. Random Fields: Rigorous Results in Statistical Mechanics and Quantum Field Theory. Ed: J. Fritz, J.L. Lebowitz, D. Szász. Elsevier North-Holland, 1981, $166 set [ISBN: 963-8021-49-7]. Volume I, 568 pp; Volume II, 540 pp. Typeset proceedings (note price!) of a June 1979 colloquium organized by the Bolyai and Bernoulli societies at Esztergom, Hungary. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

In the second game, Right colors two edges after each time Left colors one. For $n > 3$, can Right always produce a larger clique? She can for $n = 4$ or 5.

In the third game they color alternately, as in the first game, but Left wins if he can produce a vertex with larger Lilac valence than any Red valence, i.e., if he can make more Lilac edges at a vertex than Right can make Red edges at any vertex. If the maximum Lilac and Red valences are equal, Right wins. Who does win?

A Conjecture of Marian Deaconescu

Let $\Pi = p_1 p_2 \cdots p_n$ be the product of the first n primes. Then

¿ $\Pi + p$ is prime for at least one prime p, $p_n < p < \Pi$?

Erdős agrees with this conjecture and expects that the least such p is much smaller than Π; in fact that $p < n^c$ for some constant c. Deaconescu has verified the conjecture for $n \leqslant 1000$.

A PENTAD OF POINTED PROBLEMS

Quite often the MONTHLY gets problems which are too unsolved to go into the elementary or advanced sections, even with stars on, while they may be too brief to stand alone in this section. Here are five examples.

Some Problems in Geometric Probability

Paul R. Chernoff, University of California, Berkeley, CA 94720

Choose n points $P_1, P_2, \ldots, P_n$ at random from the unit disk in R^2. What is the probability that the polygon $P_1 P_2 \ldots P_n P_1$ is convex? What is the probability that it's a Jordan curve? If the points are chosen from the unit ball in R^3, what's the probability that the polygon is knotted?

I have no idea how to solve these, I believe that Dvoretzky has done some possibly related work on Brownian motion trajectories. I think that the first question may be easy, but that the last one is hard.

Find an Invertible Linear Combination

Robert Hartwig, North Carolina State University at Raleigh, NC 27650

If Q_1, Q_2, Q_3, Q_4 are square matrices of order n with elements in a field F and the matrix $\begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix}$ of order $2n$ is invertible, are there scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in F so that $U = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4$ is invertible?

What about the same problem with "field" replaced by "division ring" or "the ring of integers"? A necessary condition for matrices over a commutative ring with unity, is that the **splitting condition** holds; that $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ similar to $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$ implies A similar to B. For, if A, B are n by n matrices, then

$$\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$$

implies $AQ_i = Q_i B$ for each i, and hence since R is commutative, $AU = UB$ so that an affirmative answer to the question implies the splitting condition.

Moving Triangles in the Plane

Peter J. Giblin, University of North Carolina at Chapel Hill, NC 27514, and University of Liverpool, Liverpool, L69 3BX, England

In spite of its unnatural look, this problem arose in research into caustics by reflection, undertaken with J. W. Bruce and C. G. Gibson (*Oxford Quart. J.*, 1982):

A, B, C each move on curves in the plane. I want to impose two sets of conditions:

(1) The normal to the curve on which A moves is to be, at each instant, the bisector of angle CAB. Similarly for the normals at B and C. [Note that this implies $BC + CA + AB =$ constant (*Math. Mag.*, Problem 1151, 1982). If A, say, is permanently fixed, then the condition should be interpreted as saying that the bisector of angle CAB is permanently fixed, but in fact this is incompatible with condition (2) below.]

(2) The curvature of the curve on which A moves is to have, at each instant, the value $(\cos^3\alpha)/s$, where $2s$ is the constant perimeter noted above, and 2α is the angle CAB. Similarly the curvature of the curves on which B and C move are to be $(\cos^3\beta)/s$ and $(\cos^3\gamma)/s$, where $2\beta, 2\gamma$ are the angles ABC and BCA.

Problem. Do such moving triangles exist?

The Gauss-Lucas Hull of a Set of Complex Numbers

Lee A. Rubel, University of Illinois, Urbana, IL 61801

The Gauss-Lucas Theorem [see M. Marden, (The) Geometry of (the Zeros of) Polynomials (in a Complex Variable), Math Surveys #3, Amer. Math. Soc., 1949, pp. 14 & ff; 2nd ed., 1966, pp. 22 & ff.] states that the zeros of the derivative of a complex polynomial lie in the convex hull of the zeros of the polynomial. This problem pursues the G-L Theorem further.

If E is a subset of the complex plane, let $E^1 = \{z \in \mathbb{C}: p'(z) = 0$ for some polynomial p, all of whose zeros (possibly multiple) lie in $E\}$. For an ordinal number α, let $E^\alpha = \cup\{(E^\beta)^1: \beta < \alpha\}$. Let $E^\#$ be the E^α for the first α for which $(E^\alpha)^1 = E^\alpha$. The problem is to get a geometric description of $E^\#$ in terms of E. This seems challenging even when E is a doubleton, say $E = \{0, 1\}$, when $E^1 = \{r: 0 \leqslant r \leqslant 1, r$ rational$\}$. To see that this is really E^1, we may write any polynomial p, all of whose zeros lie in E, as $p(z) = Az^m(z-1)^n$. Then

$$\frac{p'(z)}{p(z)} = \frac{m}{z} + \frac{n}{z-1} = \frac{(m+n)z - m}{z(z-1)}$$

so that $p'(z) = 0$ precisely when $z = m/(m+n)$, as claimed.

Problem 1. What is $\{0, 1\}^2$?

Problem 2. Is $\{0, 1\}^\#$ the set of all algebraic numbers in the closed interval $[0, 1]$, all of whose algebraic conjugates also lie in $[0, 1]$?

Since it is easy to see that $\{0, 1\}^\#$ is contained in this last set, the burden is to prove or disprove the reverse containment. Many related questions suggest themselves.

Rational Roots of a Polynomial

Lee A. Rubel, University of Illinois, Urbana, IL 61801

Problem: Let p be a real polynomial of degree n such that the equation $p(x) = 0$ has n distinct rational real roots. Must there exist a real number $\varepsilon \neq 0$ such that the equation $p(x) = \varepsilon$ also has n distinct rational real roots?

I believe on other evidence that the answer is probably affirmative. It is trivially so for $n = 1$ and $n = 2$, and nontrivially so for $n = 3$—here is a proof by Richard Bishop.

Suppose that $x^3 - px + q = 0$ has three rational roots. Then $x^3 - px + q_1 = 0$ has three rational roots for q_1 running through a dense set of rational numbers in an interval which includes q (possibly at an end).

1. Let the real roots of $x^3 - px + q_1 = 0$ be r, s, t. For these roots to be real, the discriminant $-p^3/27 + q_1^2/4$ must be $\leqslant 0$ so that r, s, t exist as "functions" of q_1 for $|q_1| \leqslant \sqrt{4p^3/27}$. By the vanishing of the x^2 term, $r + s + t = 0$, $t = -r - s$. Then

$$p = -(rs + rt + st) = r^2 + rs + s^2 = \left(r + \frac{s}{2}\right)^2 + 3\left(\frac{s}{2}\right)^2.$$

The possibilities (r, s, t) as q_1 varies are parametrized by r, s on the conic $\left(r + \frac{s}{2}\right)^2 + 3\left(\frac{s}{2}\right)^2 = p$.

2. The rational values of (r, s) on the conic yield rational values of q_1 for which there are three rational roots. Thus, the result reduces to the following:

3. Let (u_0, v_0) be a rational point on $u^2 + 3v^2 = p$. Then it is easily checked that the linear transformation

$$u = u_0 y - 3v_0 z, \quad v = v_0 y + u_0 z$$

transforms the conic into $y^2 + 3z^2 = 1$ and transforms the rational points of $u^2 + 3v^2 = p$ to those of $y^2 + 3z^2 = 1$. Note that this reduces the theorem to the case $p = 1$.

4. The conic $y^2 + 3z^2 = 1$ has a rational parametrization

$$y = \frac{a^2 - 3b^2}{a^2 + 3b^2} \quad z = \frac{2ab}{a^2 + 3b^2}.$$

The remaining cases, even $n = 4$, seem difficult.

ON THE POSING OF PROBLEMS

Modeling is rightly receiving increasing attention as an important part of mathematics. A problem which arises from a practical situation usually needs formalizing before it can be attacked. In the course of this process, some features of the problem may get omitted or modified, so that the first problem may be to decide what the problem is. Different interpretations will lead to different results; it is often a matter of opinion as to which ones are interesting. It is a pity if a problem is lost because it is initially not well-posed, or if the questions are properly posed but turn out not to be the best ones to ask.

The following is one of a number of editorial paraphrases that it may be convenient to make of papers that have been submitted, or of correspondence and conversations made with referees and others. Readers may be able to ask, and answer, better questions than those which occurred to the author, the referee and the editor.

A Problem Suggested by Celtic Art

Douglas A. Engel, 2935 W. Chenango, Englewood, CO 80110

Ancient Celtic artists sometimes used the following device to produce braided patterns, although they did not exactly follow the instructions given here, which have been crystallized into a precise problem.

Draw a graph on a rectangular array of m rows of n points, where m and n are coprime and

I believe on other evidence that the answer is probably affirmative. It is trivially so for $n = 1$ and $n = 2$, and nontrivially so for $n = 3$—here is a proof by Richard Bishop.

Suppose that $x^3 - px + q = 0$ has three rational roots. Then $x^3 - px + q_1 = 0$ has three rational roots for q_1 running through a dense set of rational numbers in an interval which includes q (possibly at an end).

1. Let the real roots of $x^3 - px + q_1 = 0$ be r, s, t. For these roots to be real, the discriminant $-p^3/27 + q_1^2/4$ must be $\leqslant 0$ so that r, s, t exist as "functions" of q_1 for $|q_1| \leqslant \sqrt{4p^3/27}$. By the vanishing of the x^2 term, $r + s + t = 0$, $t = -r - s$. Then

$$p = -(rs + rt + st) = r^2 + rs + s^2 = \left(r + \frac{s}{2}\right)^2 + 3\left(\frac{s}{2}\right)^2.$$

The possibilities (r, s, t) as q_1 varies are parametrized by r, s on the conic $\left(r + \frac{s}{2}\right)^2 + 3\left(\frac{s}{2}\right)^2 = p$.

2. The rational values of (r, s) on the conic yield rational values of q_1 for which there are three rational roots. Thus, the result reduces to the following:

3. Let (u_0, v_0) be a rational point on $u^2 + 3v^2 = p$. Then it is easily checked that the linear transformation

$$u = u_0 y - 3v_0 z, \quad v = v_0 y + u_0 z$$

transforms the conic into $y^2 + 3z^2 = 1$ and transforms the rational points of $u^2 + 3v^2 = p$ to those of $y^2 + 3z^2 = 1$. Note that this reduces the theorem to the case $p = 1$.

4. The conic $y^2 + 3z^2 = 1$ has a rational parametrization

$$y = \frac{a^2 - 3b^2}{a^2 + 3b^2} \quad z = \frac{2ab}{a^2 + 3b^2}.$$

The remaining cases, even $n = 4$, seem difficult.

ON THE POSING OF PROBLEMS

Modeling is rightly receiving increasing attention as an important part of mathematics. A problem which arises from a practical situation usually needs formalizing before it can be attacked. In the course of this process, some features of the problem may get omitted or modified, so that the first problem may be to decide what the problem is. Different interpretations will lead to different results; it is often a matter of opinion as to which ones are interesting. It is a pity if a problem is lost because it is initially not well-posed, or if the questions are properly posed but turn out not to be the best ones to ask.

The following is one of a number of editorial paraphrases that it may be convenient to make of papers that have been submitted, or of correspondence and conversations made with referees and others. Readers may be able to ask, and answer, better questions than those which occurred to the author, the referee and the editor.

A Problem Suggested by Celtic Art

Douglas A. Engel, 2935 W. Chenango, Englewood, CO 80110

Ancient Celtic artists sometimes used the following device to produce braided patterns, although they did not exactly follow the instructions given here, which have been crystallized into a precise problem.

Draw a graph on a rectangular array of m rows of n points, where m and n are coprime and

just one of m and n is even. The $mn/2$ edges of the graph form a matching, or regular graph of valence one. In Figure 1(a), $m = 2$ and $n = 3$. Next arrange mn copies of the graph in n rows of m, to produce a square array of mn rows of mn points connected in pairs by $m^2n^2/2$ edges, as in Figure 1(b). Thirdly rotate this square array through 90° about its centre, as in Figure 1(c). Finally, superimpose this rotation onto the original square array to give a pattern as in Figure 1(d). This is a regular graph of valence 2 and so consists of a number of disjoint cycles. For example, the 36 edges in Figure 1(d) form a 4-cycle (shown dotted) and a 32-cycle.

The Celtic artists preferred their finished product to be a single m^2n^2-cycle or **hamilton cycle**.

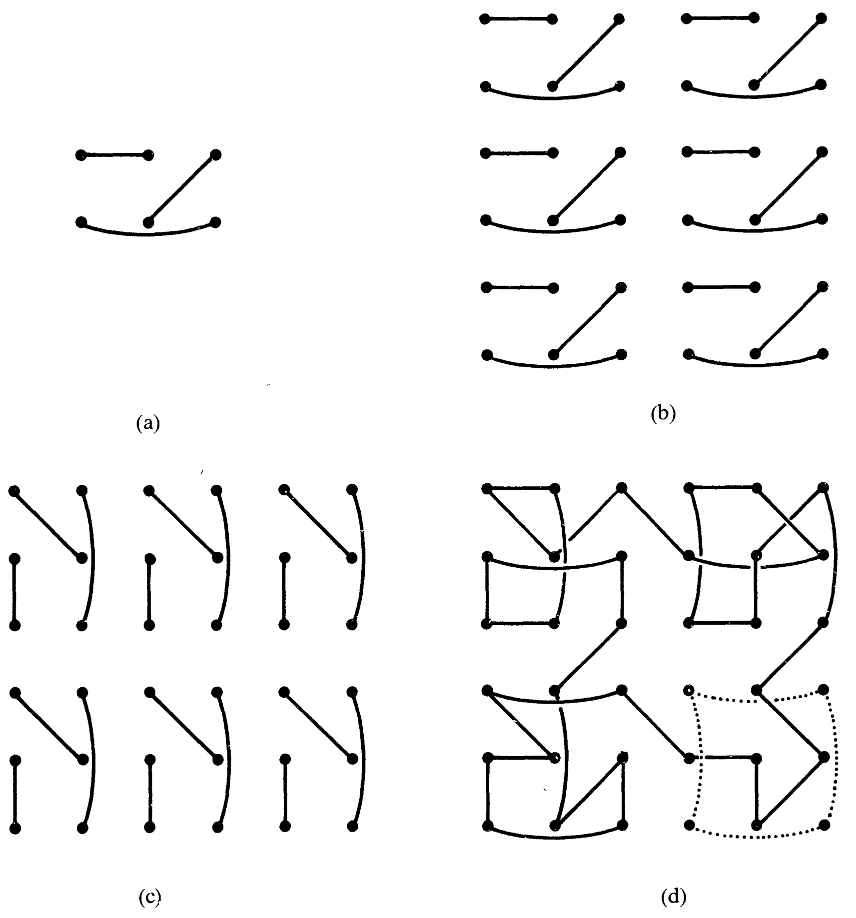


FIG. 1

Problem 1. What are the conditions on the original matching in order that the final pattern is a (single) hamilton cycle?

If $m = 1$ and $n = 2k$, then we always generate k^2 4-circuits, so that there is a hamilton cycle only in the trivial case $m = 1$, $n = 2$. Some solutions for the case $m = 3$, $n = 4$ are shown in Figure 2. [Mr. Engel has sent this editor a model of the fourth of these, made from wire and perspex; it is now on display in his Department.]

It may be possible to answer the next two problems by using a computer.

Problem 2. Find all matchings which lead to a hamilton cycle when $m = 3$ and $n = 4$.

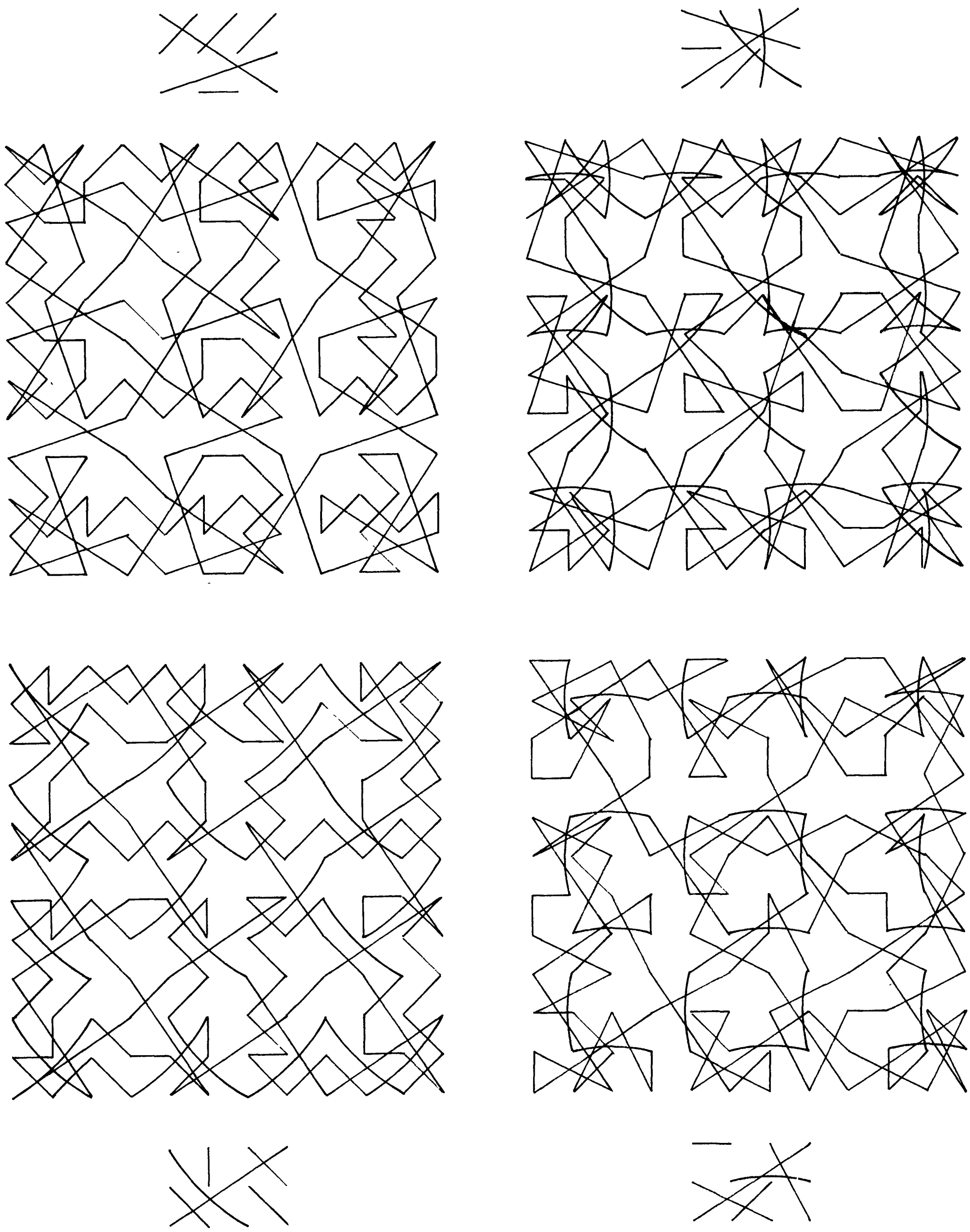


FIG. 2

Problem 3. Find such a matching when $m = 4$ and $n = 5$.

Problem 4. Are there values of m, n with $(m, n) = 1$ and mn even for which no matching generates a hamilton cycle?

As m and n increase, solutions seem harder and harder to find, but this is a common phenomenon in combinatorial searches. The number of different matchings is something like $(mn/e)^{mn/2}$ so that, even if the number of solutions increases exponentially, say like c^{mn} for some

constant c, then the chance of finding one at random is $(c^2e/mn)^{mn/2}$ which soon becomes hopelessly small as m and n increase.

Problem 5. Find bounds, or even an asymptotic formula, for the number of distinct solutions for a given m and n.

If $m = 2$ and n is odd, then solutions are known (Figure 3) which have a fourfold symmetry of rotation (through 90°).

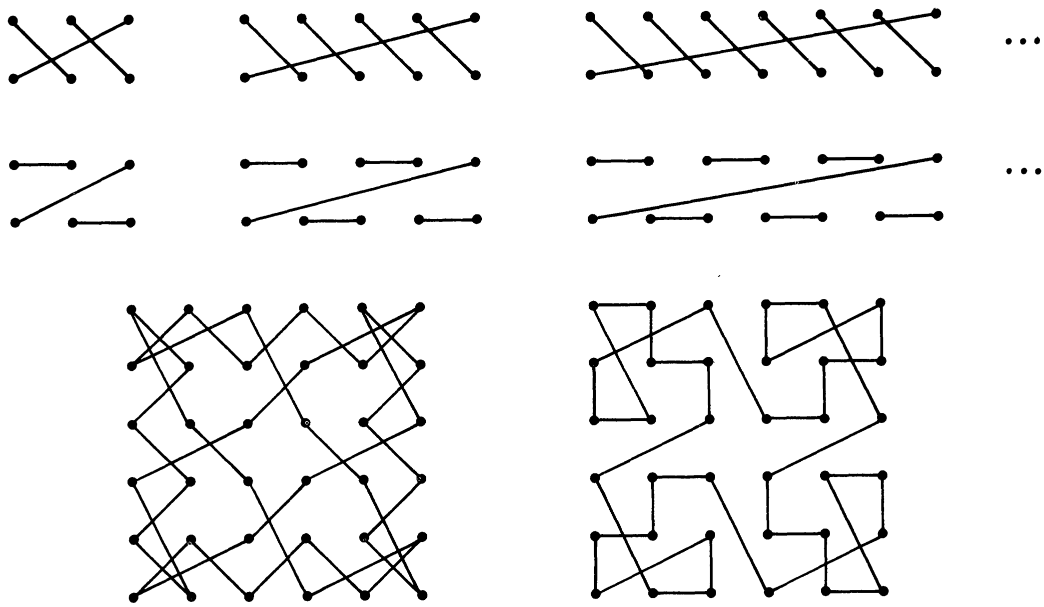


FIG. 3

Problem 6. Find such symmetrical solutions for $3 \leqslant m < n$.

Similar problems may be formulated for the case where the square array is reflected as well as rotated before being superimposed on the original. We can also ask how to formulate corresponding problems in three dimensions, based on cyrstallographic lattices.

Reference

George Bain, Celtic Art, Dover, New York, 1973.

MISCELLANEA

93. I once heard the great Richard Owen [the biologist] say... that he would like to see *Homo Mathematicus* constituted into a distinct subclass, thereby suggesting to my mind sensation, perception, reflection, abstraction, as the successive stages or phases of protoplasm on its way to being made perfect in Mathematicised Man.

–J. J. Sylvester, 1870. (Reprinted in *Collected Mathematical Papers*, vol. 2, Cambridge, 1908, p. 652.)

constant c, then the chance of finding one at random is $(c^2e/mn)^{mn/2}$ which soon becomes hopelessly small as m and n increase.

Problem 5. Find bounds, or even an asymptotic formula, for the number of distinct solutions for a given m and n.

If $m = 2$ and n is odd, then solutions are known (Figure 3) which have a fourfold symmetry of rotation (through 90°).

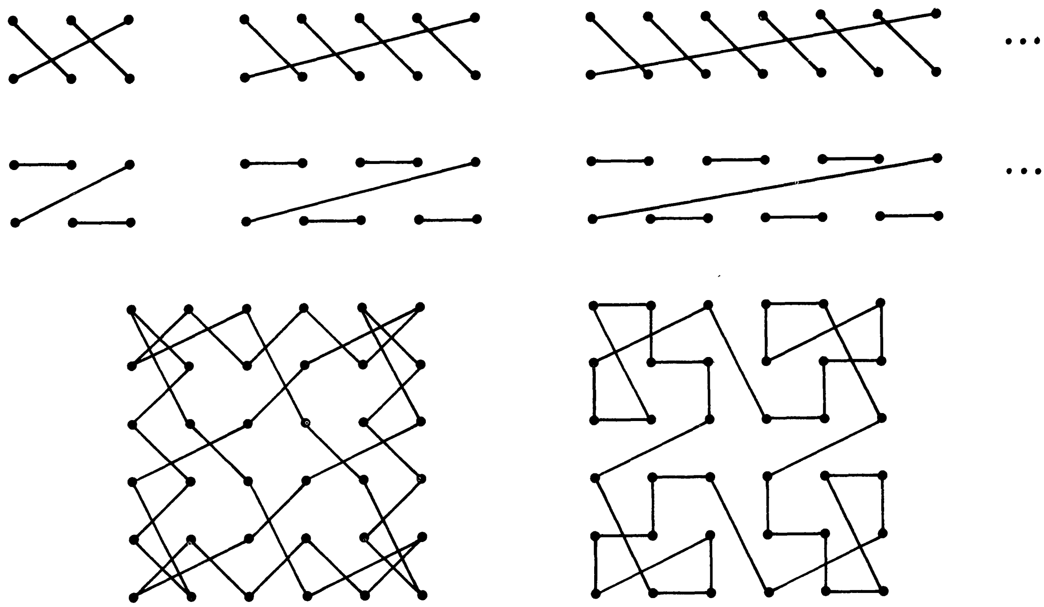


FIG. 3

Problem 6. Find such symmetrical solutions for $3 \leqslant m < n$.

Similar problems may be formulated for the case where the square array is reflected as well as rotated before being superimposed on the original. We can also ask how to formulate corresponding problems in three dimensions, based on cyrstallographic lattices.

Reference

George Bain, Celtic Art, Dover, New York, 1973.

MISCELLANEA

93. I once heard the great Richard Owen [the biologist] say... that he would like to see *Homo Mathematicus* constituted into a distinct subclass, thereby suggesting to my mind sensation, perception, reflection, abstraction, as the successive stages or phases of protoplasm on its way to being made perfect in Mathematicised Man.

–J. J. Sylvester, 1870. (Reprinted in *Collected Mathematical Papers*, vol. 2, Cambridge, 1908, p. 652.)

Now this extension is uniformly continuous on $[0, 1]$, so δ is uniformly continuous on $(0, 1]$. Now we regard δ as a function of both x and ε, and the proof is complete.

The uniformly continuous functions are not characterized by the property that they have a uniformly continuous modulus of continuity, as the following example shows.

EXAMPLE. The function $f(x) = 1/x$ on $(0, 1]$ is continuous, but not uniformly continuous. However, one may choose $\delta(x, \varepsilon) = \varepsilon x^2/2$ which is uniformly continuous.

The theorem for the Lipschitz condition is very easy to prove.

THEOREM 3. *If f: $X \to Y$ satisfies a Lipschitz condition on X, then $\delta : X \times (0, \infty) \to (0, \infty)$ may be chosen to satisfy a Lipschitz condition.*

Proof. Let $M > 0$ be a real number so that $\rho(f(x), f(y)) < M\sigma(x, y)$ for all $x, y \in X$. Then $\delta(x, \varepsilon) = \varepsilon/M$ is a modulus of continuity of f which satisfies a Lipschitz condition where the constant is $1/M$.

References

1. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.
2. S. B. Seidman and J. A. Childress, A continuous modulus of continuity, this MONTHLY, 82 (1975) 253-254.

APPLICATIONS OF A SIMPLE COUNTING TECHNIQUE

MELVIN HAUSNER
Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

In this note, we show how the simple counting technique of Theorem 1 can be applied in several different settings. Throughout, X denotes a finite set, and p denotes a prime number. For $A \subseteq X$, let $|A|$ represent the number of elements in A.

1. THEOREM. *Let f: $X \to X$ with $f^p = I$ (the identity map). Let X_0 be the set of fixed points of X. $X_0 = \{x \in X | f(x) = x\}$. Then $|X| \equiv |X_0| \bmod p$.*

Proof. For any $x \in X$, we define the orbit $\bar{x}$ of x as the set $\{x, f(x), \ldots, f^{p-1}(x)\}$. It is an easy matter to show that the orbits partition X. Clearly, $|\bar{x}| = 1$ if and only if $x = f(x)$, that is, if $x \in X_0$. Now we claim that if $|\bar{x}| > 1$, then $|\bar{x}| = p$. For if we had any duplication in $\bar{x}$, then $f^i(x) = f^j(x)$ for some $i, j, 0 \leqslant i < j < p$, so that $f^{j-i}(x) = x$. Since $f^p(x) = x$ and $(j - i, p) = 1$, it follows that $f(x) = x$, and hence $|\bar{x}| = 1$.

Finally, since there are $|X_0|$ orbits of length 1, and all other orbits, say n of them, have length p, we have $|X| = |X_0| + np$. This yields the required congruence. ∎

Probably the most familiar use of Theorem 1 is for the solution of the following problem in group theory. "If G is a group of even order, prove it has an element $a \neq e$ (the identity) satisfying $a^2 = e$." (For example, see Herstein [**1**], p. 35, problem 11.) The familiar proof pairs off x and x^{-1} when they are distinct. The remaining even number of group elements satisfy $x = x^{-1}$ or $x^2 = e$. Since e is one such element, there are an odd number of elements $x \neq e$, with $x^2 = e$. In our notation, $X = G$, $p = 2$, $f(x) = x^{-1}$, so $f^2 = I$, and $X_0 = \{x | x = x^{-1}\} = \{x | x^2 = e\}$. By Theorem 1, $|X_0| \equiv |X| \equiv 0 \bmod 2$. McKay [**3**] gives a nice direct generalization of this proof which we put into our terminology.

2. EXAMPLE (CAUCHY'S THEOREM for Groups). *Let G be a finite group, p a prime dividing $|G|$. Then G has an element of order p. Indeed the number of such elements is congruent to* $-1 \bmod p$.

Proof. Let $X =$ the set of p-tuples $(x_1, \ldots, x_p)$ where $x_i \in G$, and $x_1 x_2 \cdots x_p = e$. Note that

$|X| = |G|^{p-1} \equiv 0 \bmod p$. Since $x_1 x_2 \cdots x_p = e$ implies $x_2 \cdots x_p x_1 = e$, we may define $f(x_1, \ldots, x_p) = (x_2, \ldots, x_p, x_1)$ and we have $f\colon X \to X, f^p = I$. What is X_0? $f(x) = x$ if and only if $x = (g, g, \ldots, g)$ with $g^p = e$. Thus, $|X_0| =$ the number of $g \in X$ satisfying $g^p = e$. By Theorem 1, $|X_0| \equiv |X| \equiv 0 \bmod p$. Since $|X_0| = 1 +$ number of elements of order p, this is the result.

3. EXAMPLE (FERMAT'S THEOREM). $n^p \equiv n \bmod p$.

Proof. Consider the set X of all lattice points $(x_1, \ldots, x_p)$ with $1 \leqslant x_i \leqslant n$ and let $f(x_1, \ldots, x_p) = (x_2, \ldots, x_p, x_1)$. Clearly $f^p = I$ and $|X| = n^p$, $|X_0| = n$. Theorem 1 gives the result.

4. EXAMPLE (LUCAS' THEOREM). *Suppose*

$$n = n_0 + n_1 p + \cdots + n_k p^k;\ r = r_0 + r_1 p + \cdots + r_k p^k$$

with $0 \leqslant n_i, r_i < p$. *Then*

$$\binom{n}{r} \equiv \binom{n_0}{r_0}\binom{n_1}{r_1} \cdots \binom{n_k}{r_k} \bmod p.$$

(*Note*: $\binom{a}{b} = 0$ *if* $b > a$.)

Proof. If we write $n = Np + n_0$, $r = Rp + r_0$, where $0 \leqslant n_0, r_0 < p$, it suffices to prove

$$\binom{n}{r} \equiv \binom{N}{R}\binom{n_0}{r_0} \bmod p.$$

We define $A_i = \{(i, 1), \ldots, (i, N)\}$ for $i = 1, \ldots, p$ and $B = \{(0, 1), \ldots, (0, n_0)\}$. Then, setting

$$A = A_1 \cup \cdots \cup A_p \cup B$$

we have $|A| = Np + n_0 = n$. We now define $f\colon A \to A$ by cyclically moving the A_i's and keeping B fixed:

$$\begin{aligned} f(i, x) &= (i + 1, x) \qquad 1 \leqslant i \leqslant p - 1 \\ f(p, x) &= (1, x) \\ f(0, x) &= (0, x). \end{aligned}$$

Thus,

$$f(A_i) = A_{i+1} (1 \leqslant i \leqslant p - 1), f(A_p) = A_1, f(B) = B.$$

Clearly $f^p = I$.

We now take X as the collection of subsets $C \subseteq A$ with $|C| = r$. Since $f\colon A \to A, f$ acts naturally on subsets of A: $f(C) = \{f(x) | x \in C\}$. Since f is 1-1, $|f(C)| = |C|$, so $f\colon X \to X$, with $f^p = I$. Clearly,

$$|X| = \binom{n}{r}.$$

We now find $|X_0|$. Any subset C of A can be written uniquely as

$$C = C_1 \cup \cdots \cup C_p \cup C_0$$

where $C_i \subseteq A_i$, $C_0 \subseteq B$. Since f sends the A_i cyclically around, and keeps B fixed, we see that $f(C) = C$ if and only if

$$C_i = f^{i-1}(C_1) \qquad i = 1, \ldots, p.$$

For $C \in X$, we must have $|C| = r$, and if $C \in X_0$, we have $|C| = p|C_1| + |C_0| = r = Rp + r_0$. Note that $0 \leqslant |C_0|, r_0 < p$. Thus the cardinality restriction on C is satisfied if and only if $|C_1| = R$,

$|C_0| = r_0$. But there are $\binom{N}{R}$ such choices for C_1 and $\binom{n_0}{r_0}$ independent choices for C_0. Thus,

$$|X_0| = \binom{N}{R}\binom{n_0}{r_0}.$$

Theorem 1 then provides the desired conclusion.

References

1. I. N. Herstein, Topics in Algebra, 2nd ed., Wiley New York, 1975.
2. E. Lucas, Bull. Soc. Math. France, 6 (1877–78) 52.
3. J. H. McKay, Another Proof of Cauchy's Group Theorem, this MONTHLY, 66 (1959) 119.

THE MEANING OF THE CONJECTURE $P \neq NP$ FOR MATHEMATICAL LOGIC

JAN MYCIELSKI
Department of Mathematics University of Colorado, Boulder, Colorado 80309

I think that the conjecture $P \neq NP$ is not as widely taught in courses of mathematical logic as it should be, in view of its capital importance for the foundation of mathematics. Therefore I am writing this note in the hope that all logicians will always include it in the introductory courses of their subject although it does not appear yet in the appropriate books. The original paper of S. Cook [**1**], where the conjecture was formulated, was indeed written from the point of view of logic but it became the domain of computer scientists (see [**2**]), particularly because of a paper of Karp [**3**] where the combinatorial or computational aspects of the conjecture were developed in a very suggestive way.

Let us assume that the teacher has already presented the concept of a first order theory and the concept of a Turing machine (neither the concepts of a decidable theory nor that of a recursive function are needed). Then he may proceed as follows:

By a *normal theory* we shall mean a theory which is formalized with a finite alphabet in first order logic with equality and is axiomatizable by a finite set of axioms and axiom schemata in which one can prove $\exists xy[x \neq y]$. (In [**4**] it is proved that every theory which is recursively axiomatizable and contains a minimal amount of arithmetic or set theory is normal.) By a proof in a normal theory we mean a Hilbert style proof from the axioms.

Let Σ be a finite alphabet and Σ^* the set of all words, i.e., finite sequences of elements of Σ. For any $\xi \in \Sigma^*$, $|\xi|$ denotes the length of ξ. Now we introduce a more abstract concept of a theory which we will call a $\tau\pi$-theory.

A *$\tau\pi$-theory* is a set of pairs $T \subseteq \Sigma^* \times \Sigma^*$ such that there exists a polynomial $P(x, y)$ and a Turing machine M such that, for any $(\tau, \pi) \in \Sigma^* \times \Sigma^*$, M can decide in time $\leq P(|\tau|, |\pi|)$ if $(\tau, \pi) \in T$.

If $(\tau, \pi) \in T$, then τ is called a *theorem* of T and π is called a *proof* of τ in T.

Every normal theory defines a $\tau\pi$-theory since the time necessary to check the correctness of a Hilbert style proof in a normal theory can be estimated from above by a polynomial in the length of that proof.

Now, a $\tau\pi$-theory T will be called *amenable* (to *automatization*) iff there exists another polynomial $P_0(x, y)$ and another Turing machine M_0 such that, given any word $\tau \in \Sigma^*$ and any positive integer n, M_0 can decide in time $\leq P_0(|\tau|, n)$ if there exists a $\pi \in \Sigma^*$ with $|\pi| \leq n$ and such that $(\tau, \pi) \in T$. (Notice that if we replaced the condition $\leq P_0(|\tau|, n)$ by the condition $\leq P_0(|\tau|, c^n)$ where $c = \text{card}\,\Sigma$, then the concept would trivialize since every $\tau\pi$-theory would be amenable. In fact, given a time $P_0(|\tau|, c^n)$, the machine can form all sequences of symbols of length n and find out if any of them is a proof of τ.)

It is clear that, after Gödel's 1931 discovery that all sufficiently strong theories are undecidable,

$|C_0| = r_0$. But there are $\binom{N}{R}$ such choices for C_1 and $\binom{n_0}{r_0}$ independent choices for C_0. Thus,

$$|X_0| = \binom{N}{R}\binom{n_0}{r_0}.$$

Theorem 1 then provides the desired conclusion.

References

1. I. N. Herstein, Topics in Algebra, 2nd ed., Wiley New York, 1975.
2. E. Lucas, Bull. Soc. Math. France, 6 (1877–78) 52.
3. J. H. McKay, Another Proof of Cauchy's Group Theorem, this MONTHLY, 66 (1959) 119.

THE MEANING OF THE CONJECTURE $P \neq NP$ FOR MATHEMATICAL LOGIC

JAN MYCIELSKI

Department of Mathematics University of Colorado, Boulder, Colorado 80309

I think that the conjecture $P \neq NP$ is not as widely taught in courses of mathematical logic as it should be, in view of its capital importance for the foundation of mathematics. Therefore I am writing this note in the hope that all logicians will always include it in the introductory courses of their subject although it does not appear yet in the appropriate books. The original paper of S. Cook [**1**], where the conjecture was formulated, was indeed written from the point of view of logic but it became the domain of computer scientists (see [**2**]), particularly because of a paper of Karp [**3**] where the combinatorial or computational aspects of the conjecture were developed in a very suggestive way.

Let us assume that the teacher has already presented the concept of a first order theory and the concept of a Turing machine (neither the concepts of a decidable theory nor that of a recursive function are needed). Then he may proceed as follows:

By a *normal theory* we shall mean a theory which is formalized with a finite alphabet in first order logic with equality and is axiomatizable by a finite set of axioms and axiom schemata in which one can prove $\exists xy[x \neq y]$. (In [**4**] it is proved that every theory which is recursively axiomatizable and contains a minimal amount of arithmetic or set theory is normal.) By a proof in a normal theory we mean a Hilbert style proof from the axioms.

Let Σ be a finite alphabet and Σ^* the set of all words, i.e., finite sequences of elements of Σ. For any $\xi \in \Sigma^*$, $|\xi|$ denotes the length of ξ. Now we introduce a more abstract concept of a theory which we will call a $\tau\pi$-theory.

A *$\tau\pi$-theory* is a set of pairs $T \subseteq \Sigma^* \times \Sigma^*$ such that there exists a polynomial $P(x, y)$ and a Turing machine M such that, for any $(\tau, \pi) \in \Sigma^* \times \Sigma^*$, M can decide in time $\leq P(|\tau|, |\pi|)$ if $(\tau, \pi) \in T$.

If $(\tau, \pi) \in T$, then τ is called a *theorem* of T and π is called a *proof* of τ in T.

Every normal theory defines a $\tau\pi$-theory since the time necessary to check the correctness of a Hilbert style proof in a normal theory can be estimated from above by a polynomial in the length of that proof.

Now, a $\tau\pi$-theory T will be called *amenable* (to *automatization*) iff there exists another polynomial $P_0(x, y)$ and another Turing machine M_0 such that, given any word $\tau \in \Sigma^*$ and any positive integer n, M_0 can decide in time $\leq P_0(|\tau|, n)$ if there exists a $\pi \in \Sigma^*$ with $|\pi| \leq n$ and such that $(\tau, \pi) \in T$. (Notice that if we replaced the condition $\leq P_0(|\tau|, n)$ by the condition $\leq P_0(|\tau|, c^n)$ where $c = \text{card}\,\Sigma$, then the concept would trivialize since every $\tau\pi$-theory would be amenable. In fact, given a time $P_0(|\tau|, c^n)$, the machine can form all sequences of symbols of length n and find out if any of them is a proof of τ.)

It is clear that, after Gödel's 1931 discovery that all sufficiently strong theories are undecidable,

the next question which presents itself is to ask if the $\tau\pi$-theories (corresponding to normal theories) are amenable. But we had to wait until 1970 (the paper of Cook [**1**]) for a clear statement of that question. It can be asked in many equivalent ways and the proofs of their equivalence are very ingenious (see [**2**]). Now we will state three such ways which are most striking to the logician.

PROPOSITION. *The following three statements are equivalent to each other:*

(i) *There exists a $\tau\pi$-theory which is not amenable.*
(ii) *Every $\tau\pi$-theory which is defined by a normal theory is not amenable.*
(iii) $P \neq NP$.

This proposition follows immediately from the theorem of Cook that the set of satisfiable formulas of propositional calculus is NP-complete. (We refer the reader to [**1**], [**2**], [**3**].) It is not known if $P \neq NP$; this is the conjecture of Cook. I believe that this is the most outstanding problem of contemporary mathematics (I rank it higher than the Riemann-ζ conjecture). The above proposition is the best way of explaining the conjecture's capital importance in foundations of mathematics. (Its great importance in theoretical computer science is well known (see [**2**]).)

References

1. S. A. Cook, The Complexity of Theorem-Proving Procedures, Proc. 3rd Annual ACM Sympos. on Theory of Computing, 1970, pp. 151–158.
2. M. R. Garey and D. S. Johnson, Computers and Intractability, a Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
3. R. M. Karp, Reducibility among combinatorial problems, in the collection Complexity of Computer Computations, edited by J. W. Thatcher, Plenum Press, New York-London, 1972.
4. R. L. Vaught, Axiomatizability by a schema, J. Symbolic Logic, 32 (1967) 473–479.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

FROM CENTER OF GRAVITY TO BERNSTEIN'S THEOREM

RAY REDHEFFER
Department of Mathematics, University of California, Los Angeles CA 90024

Let $w(x)$ be continuous for $x \geqslant 0$ and positive for $x > 0$. If $w(x)$ is thought to be the density of a rod at point x, then

$$\frac{\int_0^x tw(t)\,dt}{\int_0^x w(t)\,dt}$$

is the center of gravity of the part of the rod on $[0, x]$ and, on physical grounds, this expression must be an increasing function of x. The mathematical reason is that the factor t in the numerator is increasing. In fact if ϕ is increasing for $x > 0$ and ϕw is continuous for $x \geqslant 0$, then

$$(1)\qquad \frac{\int_0^x \phi(t)w(t)\,dt}{\int_0^x w(t)\,dt} \quad \text{is increasing for } x > 0.$$

This is seen when we differentiate by the quotient rule; the numerator of the resulting expression

the next question which presents itself is to ask if the $\tau\pi$-theories (corresponding to normal theories) are amenable. But we had to wait until 1970 (the paper of Cook [**1**]) for a clear statement of that question. It can be asked in many equivalent ways and the proofs of their equivalence are very ingenious (see [**2**]). Now we will state three such ways which are most striking to the logician.

PROPOSITION. *The following three statements are equivalent to each other:*

(i) *There exists a $\tau\pi$-theory which is not amenable.*
(ii) *Every $\tau\pi$-theory which is defined by a normal theory is not amenable.*
(iii) $P \neq NP$.

This proposition follows immediately from the theorem of Cook that the set of satisfiable formulas of propositional calculus is NP-complete. (We refer the reader to [**1**], [**2**], [**3**].) It is not known if $P \neq NP$; this is the conjecture of Cook. I believe that this is the most outstanding problem of contemporary mathematics (I rank it higher than the Riemann-ζ conjecture). The above proposition is the best way of explaining the conjecture's capital importance in foundations of mathematics. (Its great importance in theoretical computer science is well known (see [**2**]).)

References

1. S. A. Cook, The Complexity of Theorem-Proving Procedures, Proc. 3rd Annual ACM Sympos. on Theory of Computing, 1970, pp. 151–158.

2. M. R. Garey and D. S. Johnson, Computers and Intractability, a Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.

3. R. M. Karp, Reducibility among combinatorial problems, in the collection Complexity of Computer Computations, edited by J. W. Thatcher, Plenum Press, New York-London, 1972.

4. R. L. Vaught, Axiomatizability by a schema, J. Symbolic Logic, 32 (1967) 473–479.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

FROM CENTER OF GRAVITY TO BERNSTEIN'S THEOREM

RAY REDHEFFER
Department of Mathematics, University of California, Los Angeles CA 90024

Let $w(x)$ be continuous for $x \geqslant 0$ and positive for $x > 0$. If $w(x)$ is thought to be the density of a rod at point x, then

$$\frac{\int_0^x tw(t)\,dt}{\int_0^x w(t)\,dt}$$

is the center of gravity of the part of the rod on $[0, x]$ and, on physical grounds, this expression must be an increasing function of x. The mathematical reason is that the factor t in the numerator is increasing. In fact if ϕ is increasing for $x > 0$ and ϕw is continuous for $x \geqslant 0$, then

$$\frac{\int_0^x \phi(t)w(t)\,dt}{\int_0^x w(t)\,dt} \quad \text{is increasing for } x > 0. \tag{1}$$

This is seen when we differentiate by the quotient rule; the numerator of the resulting expression

is

$$w(x)\left(\phi(x)\int_0^x w(t)\,dt - \int_0^x w(t)\phi(t)\,dt\right) \geqslant 0.$$

More generally, for any continuous function h let

$$J_n(h,x) = \int_0^x \int_0^{t_n} \cdots \int_0^{t_2} h(t_1)\,dt_1 \cdots dt_{n-1}dt_n, \quad n \geqslant 2$$

$$J_1(h,x) = \int_0^x h(t)\,dt.$$

Then, with w and ϕ as above,

$$\frac{J_n(\phi w, x)}{J_n(w,x)} \text{ is increasing for } x \geqslant 0. \tag{2}$$

This is true for $n = 1$ by (1). If it holds for $n - 1$ then $J_{n-1}(\phi w, x) = \Phi(x)J_{n-1}(w,x)$ where Φ is increasing. With $W(x) = J_{n-1}(w,x)$ the numerator and denominator in (2) are respectively

$$\int_0^x \Phi(t)W(t)\,dt, \quad \int_0^x W(t)\,dt.$$

Hence, (2) follows from (1) applied to (Φ, W) instead of (ϕ, w).

Suppose now that f is a real-valued function satisfying $f^{(n)}(x) \geqslant 0$ for $0 \leqslant x < \rho$, where $\rho > 0$ and $n = 0, 1, 2, \ldots$. Bernstein's theorem states that the Taylor series for f converges to $f(x)$ on $[0, \rho)$ (and hence f is the restriction of an analytic function $f(z)$, $|z| < \rho$).

To prove Bernstein's theorem, let us integrate $f^{(n)}(x)$ from 0 to x repeatedly as in [**3**]. The result is

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + \cdots + f^{(n-1)}(0)\frac{x^{n-1}}{(n-1)!} + R_n(x) \tag{3}$$

where for $n \geqslant 2$

$$R_n(x) = \int_0^x \int_0^{t_n} \cdots \int_0^{t_2} f^{(n)}(t_1)\,dt_1 \cdots dt_{n-1}dt_n = J_n(f^{(n)}, x).$$

Since $f^{(n+1)}(x) \geqslant 0$ the function $\phi = f^{(n)}$ is increasing and (2) with $w = 1$ shows that $R_n(x)/x^n$ is increasing also. Hence

$$\frac{R_n(x)}{x^n} \leqslant \frac{R_n(r)}{r^n}, \quad 0 < x \leqslant r < \rho. \tag{4}$$

The hypothesis $f^{(n)}(x) \geqslant 0$ shows that $R_n(x) \geqslant 0$ and Equation (3) with $f^{(j)}(0) \geqslant 0$, $j = 0, 1, 2, \ldots, n-1$ gives $R_n(r) \leqslant f(r)$. Thus,

$$0 \leqslant R_n(r) \leqslant f(r). \tag{5}$$

Combining (4) and (5) gives $0 \leqslant R_n(x) \leqslant (x/r)^n f(r)$, hence $R_n(x) \to 0$ for $0 \leqslant x < r$, and this is Bernstein's theorem.

In comparing with other procedures [**1**], [**2**] it should be stated that, once monotonicity of $R_n(x)/x^n$ is established, completion of the argument by (4) and (5) is common to all. The novelty here consists in the proof of monotonicity.

References

1. Tom Apostol, Mathematical Analysis, 2nd ed., Addison-Wesley, 1975, pp. 242–244.

2. Emil Artin, Calculus and Analytic Geometry, CUPM, 1957, pp. 82–90.

3. I. S. Sokolnikoff and R. M. Redheffer, Mathematics of Physics and Modern Engineering, 2nd ed., McGraw-Hill 1966, p. 36.

A NOTE ON LAGRANGE'S THEOREM

WELLS JOHNSON
Department of Mathematics, Bowdoin College, Brunswick, Maine 04011

It has become traditional in elementary abstract algebra courses to prove Lagrange's theorem (that the order of a subgroup divides the order of the group) and then to deduce from it the immediate corollary that in a finite group the order of an element always divides the order of the group. While this approach is undoubtedly the most efficient way to prove these results, we suggest below an alternative development that may be better pedagogically. Teachers do not always appreciate the fact that beginning algebra students often find the proof of Lagrange's theorem abstract and difficult. The difficulty arises from the fact that the usual proof combines two concepts which students probably have not encountered previously, namely, the idea of an equivalence relation and the notion of right or left cosets. While most students become adept at applying the result of Lagrange's theorem in different situations, only the very best ones seem to be able to reproduce its proof with any accuracy and understanding.

Our approach has been to reverse the order in the presentation of these two theorems. Many texts already define the order of an element in a group before defining the concept of a subgroup. After computing a few examples, students are easily led to the conjecture that the order of an element divides the order of the group. Historically, of course, this is a much older result than Lagrange's theorem. In number theory, the ideas go back to Fermat in the form that a prime p divides $a^p - a$ for any integer a.

This result is also easier to prove, especially if the group G is abelian. If x is a fixed element of G, then the map "left multiplication by x," $f_x : G \to G$ defined by $f_x(g) = xg$, is a permutation of the elements of G. In the case that G is abelian and has order n, then

$$\prod_{g \in G} g = \prod_{g \in G} (xg) = x^n \prod_{g \in G} g,$$

and thus $x^n = e$, the identity of G. It follows easily that the order of x divides n.

In the general nonabelian case, the proof is also easy if we are permitted to use some elementary facts about permutations. The permutation f_x defined above decomposes into disjoint cycles, each of the form $(g \ xg \ x^2g \ \ldots \ x^{k-1}g)$. The length k of a cycle exactly equals the order of x in G and hence is independent of the choice of g that begins the cycle. Thus all cycles have the same length, and the order of x times the number of distinct cycles equals the order of G, proving the result.

Having proved this theorem, we can now motivate the concepts needed for the proof of Lagrange's theorem. The idea of an equivalence relation can be introduced quite naturally by declaring two elements of G to be "equivalent" if they belong to the same cycle of the permutation f_x. It is easy to show, moreover, that two elements g and h are equivalent if and only if the product gh^{-1} is a power of x, i.e., in the cyclic subgroup generated by x. Now we are ready to generalize to arbitrary subgroups. The decomposition of f_x into disjoint cycles generalizes to the decomposition of G into disjoint right cosets and Lagrange's theorem follows easily. The point is that Lagrange's theorem is a natural generalization of the decomposition of a permutation into disjoint cycles, and students who fail to see this miss an important conceptual and historical connection.

An added dividend to this approach is that we establish very early on in the course the importance of the fact that left multiplication by x is a permutation of G. This not only aids in the construction of small group tables, but the statement and proof of Cayley's theorem are not quite the mystery to beginning students that they usually are. Once equivalence relations are understood as an abstract principle, then the development of the class equation and its generalization to group action on sets can be motivated. The latter yields some very nice proofs of Sylow's theorems.

Developed in this way, an introduction to group theory proceeds from some rather simple, concrete results to some nontrivial general theorems, following a progression of ideas that are tied very closely to one another. The presentation demonstrates how mathematicians start with basic facts that depend upon some explicit computations, and then continually rework, reformulate, and generalize the ideas, transferring them to different settings and situations. Perhaps this kind of mathematical evolution is one of the most important lessons that we can teach in an introductory abstract algebra course.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2986*. *Proposed by Abdul Aziz, University of Kashmir, India.*

If $P(z)$ is a polynomial of degree n with complex coefficients having all its zeros in $|z| \geqslant K$, where $K > 1$, prove or disprove the following two assertions:

(i) $|P(K^2z) - P(z)| \leqslant (K^n - 1)\mathrm{Max}_{|z|=1}|P(z)|$, for $|z| = 1$.
(ii) $|P(K^2z)| - |P(z)| \leqslant (K^n - 1)|z|^n\mathrm{Max}_{|z|=1}|P(z)|$, for $|z| > 1$.

E 2987. *Proposed by George Tsintsifas, Thessaloniki, Greece.*

Let $S_n = A_1A_2 \dots A_{n+1}$ be an n-simplex in $\mathbb{R}^n$ and M a point inside its circumsphere S: $(0, R)$. The straight line A_iM intersects the sphere $(0, R)$ at the point A_i'. We denote $K = \sum_{i=1}^{n+1}|A_iM|/|MA_i'|$. Let G be the centroid of S_n. Prove:

a) $K > n + 1$ iff M lies outside the sphere (s) with diameter OG.
b) $K = n + 1$ iff M lies on the sphere (s).
c) $K < n + 1$ iff M lies in the interior of the sphere (s).

Developed in this way, an introduction to group theory proceeds from some rather simple, concrete results to some nontrivial general theorems, following a progression of ideas that are tied very closely to one another. The presentation demonstrates how mathematicians start with basic facts that depend upon some explicit computations, and then continually rework, reformulate, and generalize the ideas, transferring them to different settings and situations. Perhaps this kind of mathematical evolution is one of the most important lessons that we can teach in an introductory abstract algebra course.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

Send all ***proposed*** *problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2986*. *Proposed by Abdul Aziz, University of Kashmir, India.*

If $P(z)$ is a polynomial of degree n with complex coefficients having all its zeros in $|z| \geqslant K$, where $K > 1$, prove or disprove the following two assertions:

(i) $|P(K^2z) - P(z)| \leqslant (K^n - 1)\mathrm{Max}_{|z|=1}|P(z)|$, for $|z| = 1$.
(ii) $|P(K^2z)| - |P(z)| \leqslant (K^n - 1)|z|^n\mathrm{Max}_{|z|=1}|P(z)|$, for $|z| > 1$.

E 2987. *Proposed by George Tsintsifas, Thessaloniki, Greece.*

Let $S_n = A_1A_2 \dots A_{n+1}$ be an n-simplex in $\mathbb{R}^n$ and M a point inside its circumsphere S: $(0, R)$. The straight line A_iM intersects the sphere $(0, R)$ at the point A_i'. We denote $K = \sum_{i=1}^{n+1}|A_iM|/|MA_i'|$. Let G be the centroid of S_n. Prove:

a) $K > n + 1$ iff M lies outside the sphere (s) with diameter OG.
b) $K = n + 1$ iff M lies on the sphere (s).
c) $K < n + 1$ iff M lies in the interior of the sphere (s).

E 2988. *Proposed by Mihaly Bencze, Braga, Romania.*

For what values of n do the inequalities

$$n + \ln\frac{n+2}{3} < 1 + \sqrt{2} + \sqrt[3]{3} + \cdots + \sqrt[n]{n} < 2n - \ln(n+1)$$

hold?

SOLUTIONS OF ELEMENTARY PROBLEMS

kth Divided Difference of a Function

E 2900 [1981, 538]. *Proposed by Edward Neuman, University of Wroclaw, Poland.*

The kth divided difference of a function f at the points $t_0, t_1, \ldots, t_k$ is denoted by $[t_0, t_1, \ldots, t_k]f$ and is defined inductively by

$$[t_0]f = f(t_0), [t_0, t_1, \ldots, t_{k+1}]f = ([t_1, t_2, \ldots, t_{k+1}]f - [t_0, t_1, \ldots, t_k]f)/(t_{k+1} - t_0).$$

Show that, for positive integers k, l,

$$[t_0, t_1, \ldots, t_k]x^{k+l} = \sum t_{j1}t_{j2} \cdots t_{jl},$$

the sum being extended over $0 \leqslant j_1 \leqslant j_2 \leqslant \cdots \leqslant j_l \leqslant k$.

Solution by Miroslav D. Ašić, London School of Economics. A solution to this problem can be found in [**2**], pp. 7–8, where it is solved by putting a conveniently chosen rational function into partial fractions. Here we give another proof based on the well-known formula [**2**, p. 10] for divided differences: $[t_0, \ldots, t_k]f = D(f)/d$, where $d = \Pi_{i<j} (t_i - t_j)$ and $D(f) = \det A_{(k+1)\times(k+1)}$, where A is the matrix with rth column $[f(t_r), t_r^{k-1}, t_r^{k-2}, \ldots, 1]^*$ $(0 \leqslant r \leqslant k)$. In our case we have $f(t) = t^{k+l}$. The determinant can be evaluated in much the same way as the Vandermonde determinant. For other methods of evaluating it, see [**1**, example 9, p. 27] or [**3**, §335, pp. 329–331].

References

1. W. L. Ferrar, Algebra, Oxford 1941.
2. L. M. Milne-Thomson: The calculus of finite differences, Macmillan, London, 1951.
3. T. Muir, Determinants, Dover, 1960.

Also solved by V. Anantharam, K. L. Bernstein, P. S. Bruckman, L. Kuipers (Switzerland), O. P. Lossers (Netherlands), W. A. Newcomb, B. N. Parlett, M. R. Railkar (India), R. E. Shafer, J. M. Stark, P. -Y. Wu (Taiwan), an anonymous solver, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by June 30, 1983. The solver's full post-office address should be on each sheet.

6421. *Proposed by J. Beck, F. Galvin, and J. Pach, Hungarian Academy of Sciences and University of Kansas.*

Let E be a set of points in the plane with the property that every closed disc of radius 1 includes at least one element of E. Prove that there exists a straight line L such that the orthogonal projection of E onto L is everywhere dense in L.

E 2988. *Proposed by Mihaly Bencze, Braga, Romania.*

For what values of n do the inequalities

$$n + \ln\frac{n+2}{3} < 1 + \sqrt{2} + \sqrt[3]{3} + \cdots + \sqrt[n]{n} < 2n - \ln(n+1)$$

hold?

SOLUTIONS OF ELEMENTARY PROBLEMS

kth Divided Difference of a Function

E 2900 [1981, 538]. *Proposed by Edward Neuman, University of Wroclaw, Poland.*

The kth divided difference of a function f at the points $t_0, t_1, \ldots, t_k$ is denoted by $[t_0, t_1, \ldots, t_k]f$ and is defined inductively by

$$[t_0]f = f(t_0), [t_0, t_1, \ldots, t_{k+1}]f = ([t_1, t_2, \ldots, t_{k+1}]f - [t_0, t_1, \ldots, t_k]f)/(t_{k+1} - t_0).$$

Show that, for positive integers k, l,

$$[t_0, t_1, \ldots, t_k]x^{k+l} = \sum t_{j1}t_{j2} \cdots t_{jl},$$

the sum being extended over $0 \leqslant j_1 \leqslant j_2 \leqslant \cdots \leqslant j_l \leqslant k$.

Solution by Miroslav D. Ašić, London School of Economics. A solution to this problem can be found in [**2**], pp. 7–8, where it is solved by putting a conveniently chosen rational function into partial fractions. Here we give another proof based on the well-known formula [**2**, p. 10] for divided differences: $[t_0, \ldots, t_k]f = D(f)/d$, where $d = \Pi_{i<j}\ (t_i - t_j)$ and $D(f) = \det A_{(k+1)\times(k+1)}$, where A is the matrix with rth column $[f(t_r), t_r^{k-1}, t_r^{k-2}, \ldots, 1]^*$ $(0 \leqslant r \leqslant k)$. In our case we have $f(t) = t^{k+l}$. The determinant can be evaluated in much the same way as the Vandermonde determinant. For other methods of evaluating it, see [**1**, example 9, p. 27] or [**3**, §335, pp. 329–331].

References

1. W. L. Ferrar, Algebra, Oxford 1941.

2. L. M. Milne-Thomson: The calculus of finite differences, Macmillan, London, 1951.

3. T. Muir, Determinants, Dover, 1960.

Also solved by V. Anantharam, K. L. Bernstein, P. S. Bruckman, L. Kuipers (Switzerland), O. P. Lossers (Netherlands), W. A. Newcomb, B. N. Parlett, M. R. Railkar (India), R. E. Shafer, J. M. Stark, P. -Y. Wu (Taiwan), an anonymous solver, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by June 30, 1983. The solver's full post-office address should be on each sheet.

6421. *Proposed by J. Beck, F. Galvin, and J. Pach, Hungarian Academy of Sciences and University of Kansas.*

Let E be a set of points in the plane with the property that every closed disc of radius 1 includes at least one element of E. Prove that there exists a straight line L such that the orthogonal projection of E onto L is everywhere dense in L.

E 2988. *Proposed by Mihaly Bencze, Braga, Romania.*

For what values of n do the inequalities

$$n + \ln\frac{n+2}{3} < 1 + \sqrt{2} + \sqrt[3]{3} + \cdots + \sqrt[n]{n} < 2n - \ln(n+1)$$

hold?

SOLUTIONS OF ELEMENTARY PROBLEMS

kth Divided Difference of a Function

E 2900 [1981, 538]. *Proposed by Edward Neuman, University of Wroclaw, Poland.*

The kth divided difference of a function f at the points $t_0, t_1, \ldots, t_k$ is denoted by $[t_0, t_1, \ldots, t_k]f$ and is defined inductively by

$$[t_0]f = f(t_0), [t_0, t_1, \ldots, t_{k+1}]f = ([t_1, t_2, \ldots, t_{k+1}]f - [t_0, t_1, \ldots, t_k]f)/(t_{k+1} - t_0).$$

Show that, for positive integers k, l,

$$[t_0, t_1, \ldots, t_k]x^{k+l} = \sum t_{j1}t_{j2} \cdots t_{jl},$$

the sum being extended over $0 \leqslant j_1 \leqslant j_2 \leqslant \cdots \leqslant j_l \leqslant k$.

Solution by Miroslav D. Ašić, London School of Economics. A solution to this problem can be found in [**2**], pp. 7–8, where it is solved by putting a conveniently chosen rational function into partial fractions. Here we give another proof based on the well-known formula [**2**, p. 10] for divided differences: $[t_0, \ldots, t_k]f = D(f)/d$, where $d = \Pi_{i<j}\ (t_i - t_j)$ and $D(f) = \det A_{(k+1)\times(k+1)}$, where A is the matrix with rth column $[f(t_r), t_r^{k-1}, t_r^{k-2}, \ldots, 1]^*$ $(0 \leqslant r \leqslant k)$. In our case we have $f(t) = t^{k+l}$. The determinant can be evaluated in much the same way as the Vandermonde determinant. For other methods of evaluating it, see [**1**, example 9, p. 27] or [**3**, §335, pp. 329–331].

References

1. W. L. Ferrar, Algebra, Oxford 1941.

2. L. M. Milne-Thomson: The calculus of finite differences, Macmillan, London, 1951.

3. T. Muir, Determinants, Dover, 1960.

Also solved by V. Anantharam, K. L. Bernstein, P. S. Bruckman, L. Kuipers (Switzerland), O. P. Lossers (Netherlands), W. A. Newcomb, B. N. Parlett, M. R. Railkar (India), R. E. Shafer, J. M. Stark, P. -Y. Wu (Taiwan), an anonymous solver, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by June 30, 1983. The solver's full post-office address should be on each sheet.

6421. *Proposed by J. Beck, F. Galvin, and J. Pach, Hungarian Academy of Sciences and University of Kansas.*

Let E be a set of points in the plane with the property that every closed disc of radius 1 includes at least one element of E. Prove that there exists a straight line L such that the orthogonal projection of E onto L is everywhere dense in L.

SOLUTIONS OF ADVANCED PROBLEMS

Dense Chains

5540 [1967, 1269]. *Proposed by G. F. Schumm, University of Chicago.*

If X and Y are chains, then X is said to be *dense* in Y if for every $A, B \in Y, A \neq B$, there exists $F \in X$ such that $A \subset F \subset B$ (where $\subset$ denotes a proper subset). Prove or disprove that for an uncountable chain Z there exists a countable chain dense in it if and only if there exists a countable set T such that $(B - A) \cap T \neq \varnothing$ for all $A, B \in Z, A \subset B$. Does there exist for every cardinal number $\aleph$, a chain of cardinality $2^{\aleph}$ having a chain of cardinality $\aleph$ dense in it?

Solution by M. Machover, St. John's University, New York. In the first part the implication does not hold both ways. For $Z = \{(-\infty, r) | r \in R\} \cup \{(-\infty, 0]\}$ is an uncountable chain of intervals ordered by set inclusion and $T = Q$ (the rationals) has the required property. Yet no chain X can be dense in Z for there is no M such that $(-\infty, 0) \subset M \subset (-\infty, 0]$.

With an additional hypothesis we can make the bi-implication true and prove a more general theorem. By a jump in a chain Z we mean an ordered pair (A, B) with $A, B \in Z, A \subset B$, and with no set of Z strictly between A and B. Denseness of X in Y is interpreted so as not to require that Y contain X as a subchain since otherwise any chain Z with a jump would be a trivial counterexample.

THEOREM 1. *A chain Z of cardinality greater than the infinite cardinal α has a dense chain of cardinality α if and only if,*

(1) *There exists a set T,* card $(T) \leqslant \alpha$, *such that $(B \setminus A) \cap T \neq \varnothing$ for all $A, B \in Z, A \subset B$;*
(2) *Z contains no singleton jumps,* i.e., *for every jump (A, B) in Z,* card $(B \setminus A) \geqslant 2$.

Proof. First, suppose Z has a dense chain X of cardinality α. Then, as the above illustration shows, Z does not contain any singleton jumps. Let $X_1 = \{N \setminus M | M, N \in X, M \subset N\}$ which has cardinality α and $\varnothing \in X_1$. Let T_1 be a choice set for X_1, that is, from each pair $M, N \in X, M \subset N$ choose a $t \in N \setminus M$ (the same t may be chosen more than once). Then card$(T_1) \leqslant \alpha$.

If Z has no jumps, define $T_2 = \varnothing$; otherwise, let J be the set of jumps in Z. Clearly, if $(A_1, B_1), (A_2, B_2)$ are distinct jumps, then either $B_1 \subseteq A_2$ or $B_2 \subseteq A_1$. It follows that card$(J) \leqslant \alpha$ since there is an $M \in X$ strictly between A and B for every jump (A, B). For each such jump choose $t \in B \setminus A$ thereby obtaining a set T_2, card$(T_2) \leqslant \alpha$, and set finally $T = T_1 \cup T_2$. Then card$(T) \leqslant \alpha$ and if $A, B \in Z, A \subset B$ and (A, B) is a jump it follows $T \cap (B \setminus A) \supseteq T_2 \cap (B \setminus A) \neq \varnothing$. Otherwise $\exists D \in Z$ and $M, N \in X$ such that $A \subset M \subset D \subset N \subset B$, so that $T \cap (B \setminus A) \supseteq T_1 \cap (B \setminus A) \neq \varnothing$.

Conversely, suppose properties (1) and (2) hold. Let $T_1 = \{t \in T | \exists A, B \in Z, A \subset B, t \in B \setminus A\}$, so that card$(T_1) \leqslant \alpha$. For $t \in T_1$ define the nonempty sets

$$A_1(t) = \{A \in Z | \exists B \in Z, A \subset B, t \in B \setminus A\},$$
$$B_1(t) = \{B \in Z | \exists A \in Z, A \subset B, t \in B \setminus A\}$$

and define functions h, g on T_1 by,

$$h(t) = \bigcup \{A | A \in A_1(t)\}, \quad g(t) = \bigcap \{B | B \in B_1(t)\}, \quad t \in T_1.$$

If J is the set of jumps in Z, then card$(J) \leqslant \alpha$ because the map that assigns to $(A, B) \in J$ any element in $(B \setminus A) \cap T$ is $(1, 1)$ with range in T. For every $(A, B) \in J, B \setminus A$ has at least two elements, say b_1, b_2, and we define a map k on J by setting in every case $k(A, B) = A \cup \{b_1\}$.

We assert that $X_1 = h(T_1) \cup g(T_1) \cup k(J)$ is a chain dense in Z. For, an analysis by cases, using in every case the linear ordering of Z, shows that X is linearly ordered by inclusion. As for

denseness let $A, B \in Z$, $A \subset B$ and choose any $t \in T \cap (B \setminus A)$. Thus $t \in T_1$ and $A \subseteq h(t) \subseteq g(t) \subseteq B$. If $h(t) \notin Z$, take $M = h(t)$ and if $h(t) \in Z$ but $g(t) \notin Z$, take $M = g(t)$. Otherwise $(h(t), g(t)) \in J$, in which case take $M = k(h(t), g(t))$. In any event $M \in X_1$ and $A \subset M \subset B$.

Finally, since $\mathrm{card}(X_1) \leqslant \alpha$, we can obtain a dense chain X of cardinality α by adding to X_1 an arbitrary subset of Z of cardinality α. Q.E.D.

The first part of the problem is solved by taking $\alpha = \aleph_0$ in the above theorem, noting that in property (1) of necessity $\mathrm{card}(T) = \aleph_0$. The second part is answered by the following theorem, at least for Zermelo-Skolem Fraenkel set theory with the axion of choice (ZFC).

THEOREM 2. *Let π be the proposition that for every cardinal number $\aleph$ there exists a chain of cardinality $2^{\aleph}$ having a chain of cardinality $\aleph$ dense in it. If* GCH *denotes the generalized continuum hypothesis, then in* ZFC, GCH $\Rightarrow$ π *but* π *is not equivalent to* GCH.

Proof. Sierpinski proved [W. Sierpinski, Sur un problème concernant les sous-ensembles croissants du continu, Fund. Math., 3(1922) 109-112] that for every $\aleph$ there exists a chain of cardinality greater than $\aleph$ containing a dense chain of cardinality $\aleph$, so that from an assumption of GCH it would follow that π was true. In fact, regarding $\aleph$ in the usual way as an ordinal and an ordinal as equal to the set of all smaller ordinals, a slight modification of Sierpinski's proof yields (assuming GCH) the following example. Let $Z_0 = \{f | f\colon \aleph \to \{0,1\}\}$ and totally order Z_0 by $f_1 \leqslant f_2$ if and only if $f_1 = f_2$ or $f_1(\alpha) < f_2(\alpha)$ at least ordinal α where they differ. Let $X = \{f \in Z_0 | \exists \alpha < \aleph, f(\beta) = 0 \text{ for } \alpha \leqslant \beta < \aleph\}$, $Y = \{f \in Z_0 | \exists \alpha < \aleph, f(\beta) = 1 \text{ for } \alpha \leqslant \beta < \aleph\}$ and set $Z = Z_0 \setminus Y$. Then $\mathrm{card}(Z) = 2^{\aleph}$, $\mathrm{card}(X) = \aleph$, and $X \subseteq Z$ is dense in Z. To pass from chains of functions to chains of sets use the order-isomorphism $f \to \{g \in Z | g \leqslant f\}$.

On the other hand W. Mitchell proved [W. Mitchell, Aronszajn Trees and the Independence of the Transfer Property, Annals of Mathematical Logic 5(1972) No. 1, 21–46] that the assertion π^*: "if S is any infinite set, then $\mathcal{P}(S)$ (power set) contains a chain of cardinality $2^{\mathrm{card}(S)}$" is in ZFC implied by GCH but is not equivalent to it. So to complete the proof we need only show that $\pi \Leftrightarrow \pi^*$ in ZFC. Given π^* let $\mathrm{card}(S) = \aleph$ and Z_0 be a chain in $\mathcal{P}(S)$ of cardinality $2^{\aleph}$. Deleting from Z_0 all sets B which occur in singleton jumps (A, B) we obtain a chain Z of cardinality $2^{\aleph}$ without singleton jumps and it follows from Theorem 1 with $T = S$ that Z has a dense chain of cardinality $\aleph$.

Conversely, given π let $\mathrm{card}(S) = \aleph$ and Z be any chain of sets of cardinality $2^{\aleph}$ having a dense chain X of cardinality $\aleph$. Let $h\colon X \to S$ be a (1, 1) correspondence and define $g\colon Z \to \mathcal{P}(S)$ by $g(A) = \{x \in S | h^{-1}(x) \subseteq A\}$, $A \in Z$. Then $g(Z)$ is a chain of subsets of S having cardinality $2^{\aleph}$.

Borel Subsets of a Product Space

6023* [1975, 308]. *Proposed by S. J. Sidney, University of Connecticut.*

If for each k in the uncountable index set K, I_k denotes a copy of $[0,1]$ and U_k denotes the copy of $(0,1]$ contained therein, prove or disprove that $\prod_k U_k$ is a Borel set in the compact space $\prod_k I_k$.

Solution by M. J. Pelling, Balliol College, Oxford. Let us generalize this problem and consider $Q = \prod_k U_k$ where each U_k is a proper Borel subset of I_k of Lebesgue measure 1. We prove that Q is not a Borel set in $X = \prod_k I_k$ by constructing a regular Borel measure μ on X and showing that Q is not measurable. This will depend on showing that the measure of open and closed sets in X can be defined in a particularly simple way.

Construction of μ. Let B be the collection of all sets $J_1 \times L' \subseteq X$ where J_1 is any open subset of any copy $L = \prod_{i=1}^{n} I_{k_i}$ of $[0,1]^n$, $n \geqslant 1$, and $L' = \prod_{k \neq k_i} I_k$. B is a base for the open sets of X. Lebesgue measure in $[0,1]^n$ or its copy L will be denoted m (n takes all values $1, 2, \ldots$ but context will clarify any ambiguity in m). For sets $J \in B$ define $\mu(J) = m(J_1)$ if $J = J_1 \times L'$.

denseness let $A, B \in Z, A \subset B$ and choose any $t \in T \cap (B \setminus A)$. Thus $t \in T_1$ and $A \subseteq h(t) \subseteq g(t) \subseteq B$. If $h(t) \notin Z$, take $M = h(t)$ and if $h(t) \in Z$ but $g(t) \notin Z$, take $M = g(t)$. Otherwise $(h(t), g(t)) \in J$, in which case take $M = k(h(t), g(t))$. In any event $M \in X_1$ and $A \subset M \subset B$.

Finally, since $\operatorname{card}(X_1) \leqslant \alpha$, we can obtain a dense chain X of cardinality α by adding to X_1 an arbitrary subset of Z of cardinality α. Q.E.D.

The first part of the problem is solved by taking $\alpha = \aleph_0$ in the above theorem, noting that in property (1) of necessity $\operatorname{card}(T) = \aleph_0$. The second part is answered by the following theorem, at least for Zermelo-Skolem Fraenkel set theory with the axion of choice (ZFC).

THEOREM 2. *Let π be the proposition that for every cardinal number $\aleph$ there exists a chain of cardinality $2^{\aleph}$ having a chain of cardinality $\aleph$ dense in it. If* GCH *denotes the generalized continuum hypothesis, then in* ZFC, GCH $\Rightarrow \pi$ *but π is not equivalent to* GCH.

Proof. Sierpinski proved [W. Sierpinski, Sur un problème concernant les sous-ensembles croissants du continu, Fund. Math., 3(1922) 109-112] that for every $\aleph$ there exists a chain of cardinality greater than $\aleph$ containing a dense chain of cardinality $\aleph$, so that from an assumption of GCH it would follow that π was true. In fact, regarding $\aleph$ in the usual way as an ordinal and an ordinal as equal to the set of all smaller ordinals, a slight modification of Sierpinski's proof yields (assuming GCH) the following example. Let $Z_0 = \{f | f: \aleph \to \{0,1\}\}$ and totally order Z_0 by $f_1 \leqslant f_2$ if and only if $f_1 = f_2$ or $f_1(\alpha) < f_2(\alpha)$ at least ordinal α where they differ. Let $X = \{f \in Z_0 | \exists \alpha < \aleph, f(\beta) = 0 \text{ for } \alpha \leqslant \beta < \aleph\}$, $Y = \{f \in Z_0 | \exists \alpha < \aleph, f(\beta) = 1 \text{ for } \alpha \leqslant \beta < \aleph\}$ and set $Z = Z_0 \setminus Y$. Then $\operatorname{card}(Z) = 2^{\aleph}$, $\operatorname{card}(X) = \aleph$, and $X \subseteq Z$ is dense in Z. To pass from chains of functions to chains of sets use the order-isomorphism $f \to \{g \in Z | g \leqslant f\}$.

On the other hand W. Mitchell proved [W. Mitchell, Aronszajn Trees and the Independence of the Transfer Property, Annals of Mathematical Logic 5(1972) No. 1, 21–46] that the assertion π^*: "if S is any infinite set, then $\mathcal{P}(S)$ (power set) contains a chain of cardinality $2^{\operatorname{card}(S)}$" is in ZFC implied by GCH but is not equivalent to it. So to complete the proof we need only show that $\pi \Leftrightarrow \pi^*$ in ZFC. Given π^* let $\operatorname{card}(S) = \aleph$ and Z_0 be a chain in $\mathcal{P}(S)$ of cardinality $2^{\aleph}$. Deleting from Z_0 all sets B which occur in singleton jumps (A, B) we obtain a chain Z of cardinality $2^{\aleph}$ without singleton jumps and it follows from Theorem 1 with $T = S$ that Z has a dense chain of cardinality $\aleph$.

Conversely, given π let $\operatorname{card}(S) = \aleph$ and Z be any chain of sets of cardinality $2^{\aleph}$ having a dense chain X of cardinality $\aleph$. Let $h: X \to S$ be a $(1,1)$ correspondence and define $g: Z \to \mathcal{P}(S)$ by $g(A) = \{x \in S | h^{-1}(x) \subseteq A\}$, $A \in Z$. Then $g(Z)$ is a chain of subsets of S having cardinality $2^{\aleph}$.

Borel Subsets of a Product Space

6023* [1975, 308]. *Proposed by S. J. Sidney, University of Connecticut.*

If for each k in the uncountable index set K, I_k denotes a copy of $[0,1]$ and U_k denotes the copy of $(0,1]$ contained therein, prove or disprove that $\prod_k U_k$ is a Borel set in the compact space $\prod_k I_k$.

Solution by M. J. Pelling, Balliol College, Oxford. Let us generalize this problem and consider $Q = \prod_k U_k$ where each U_k is a proper Borel subset of I_k of Lebesgue measure 1. We prove that Q is not a Borel set in $X = \prod_k I_k$ by constructing a regular Borel measure μ on X and showing that Q is not measurable. This will depend on showing that the measure of open and closed sets in X can be defined in a particularly simple way.

Construction of μ. Let B be the collection of all sets $J_1 \times L' \subseteq X$ where J_1 is any open subset of any copy $L = \prod_{i=1}^{n} I_{k_i}$ of $[0,1]^n$, $n \geqslant 1$, and $L' = \prod_{k \neq k_i} I_k$. B is a base for the open sets of X. Lebesgue measure in $[0,1]^n$ or its copy L will be denoted m (n takes all values $1, 2, \ldots$ but context will clarify any ambiguity in m). For sets $J \in B$ define $\mu(J) = m(J_1)$ if $J = J_1 \times L'$.

To extend μ to a regular Borel measure on X, we employ the theory of p. 231f in Halmos' *Measure Theory* (D. Van Nostrand, 1950), referred to hereafter as HMT. On the compact (= closed as X is compact Hausdorff) subsets of X define

$$c(F)=\inf_{F\subseteq J\in B}\mu(J),\ F\text{ compact}.$$

To show that c is a content (HMT p. 231) for X, we verify Halmos' defining conditions (a)–(d). It is obvious that

(a) $$0\leqslant c(F)\leqslant 1<\infty,$$

(b) $$F_1\subseteq F_2\Rightarrow c(F_1)\leqslant c(F_2).$$

(c) If F_1 and F_2 are disjoint, then since X is normal and compact and B is a base for the open sets closed under finite unions, there are sets J_1, J_2 in B such that $F_1\subseteq J_1$, $F_2\subseteq J_2$ and $J_1\cap J_2=\varnothing$. Since $\mu(J)\leqslant\mu(J')$, if $J\subseteq J'$ and $\mu(J)+\mu(J')\geqslant\mu(J\cup J')$ with equality if J, J' are disjoint, for any J, J' in B, it follows readily that

$$c(F_1\cup F_2)=c(F_1)+c(F_2).$$

(d) If F_1 and F_2 are not necessarily disjoint, then from the properties of μ stated in (c) we have

$$c(F_1\cup F_2)\leqslant c(F_1)+c(F_2).$$

Hence c is a content for X.

LEMMA. *The content c is regular* (HMT p. 237).

Proof. Given F compact and $\varepsilon>0$, suppose $F\subseteq J_1\times L'\in B$ and $m(J_1)\leqslant c(F)+\varepsilon$. If F^* is the projection of F into L, then F^* is compact, $F^*\subseteq J_1$ and $F\subseteq F^*\times L'$. Now m, regarded as a content in L, is certainly regular so there is a compact $F_1\subseteq L$ such that $F^*\subseteq \operatorname{Int} F_1\subseteq F_1$ and $m(F_1\setminus F^*)<\varepsilon$. If $D=F_1\times L'$, then D is compact, $\operatorname{Int} D=(\operatorname{Int} F_1)\times L'$, $F\subseteq\operatorname{Int} D\subseteq D$ and $c(D)=mF_1<mF^*+\varepsilon\leqslant mJ_1+\varepsilon\leqslant c(F)+2\varepsilon$. Hence,

$$c(F)=\inf\{c(D)\mid F\subseteq\operatorname{Int} D\subseteq D\text{ compact}\}$$

so that by definition c is regular. Q.E.D.

It follows from HMT p. 237, Theorem A, that c induces a regular Borel measure on X, μ say, such that for compact F, $c(F)=\mu(F)$. Our notation is permissible, i.e., this measure agrees with the μ originally defined on B, since given $J_1\times L'\in B$,

$$mJ_1=1-m(L\setminus J_1)=1-c(L\setminus J_1\times L')=1-\mu(X\setminus(J_1\times L')).$$

All the above holds if K is countably infinite but since then X has a countable base for its open sets it is easily shown that μ is actually the product measure on X, each I_k being assigned the Borel measure obtained by restricting Lebesgue measure. In this case we denote μ by μ_∞. When K is uncountable, μ properly extends the product measure.

The measure of open and closed sets. In the course of proving the lemma above it was shown incidentally that for compact F

$$\mu(F)=\inf_{F\subseteq F_1\times L'}\mu(F_1\times L')\quad\text{where } F_1 \text{ is a closed subset of } L.$$

Taking complements it follows that for any open G in X,

$$\mu(G)=\sup_{J_1\times L'\subseteq G}\mu(J_1\times L')\quad\text{where } J_1 \text{ is an open subset of } L.$$

Take a sequence $J_i\times L_i'\subseteq G$ such that $\mu(J_i\times L_i')\to\mu(G)$. Then

$$\bigcup_{i=1}^{\infty}J_i\times L_i'=T\times M'\subseteq G$$

where T is an open set in some countably infinite product $M = \prod_{j=1}^{\infty} I_{k_j}$ and $M' = \prod_{k \neq k_j} I_k$. Further $\mu(G) = \mu(T \times M') = \mu_\infty(T)$.

If now $G = \cup_\alpha J_\alpha$, $J_\alpha \in B$, let S_α be the projection of J_α into M so $S = \cup_\alpha S_\alpha$ will be the projection of G into M. S and the S_α are all open in M so that as M has a countable base for its open sets, there is a countable collection $S_1, S_2, \ldots$ of the S_α such that $S = \cup_{i=1}^{\infty} S_i$. Each $\mu_\infty(S_i \setminus T) = 0$ else we would have $\mu(G) > \mu(T \times M')$ and so

$$\mu_\infty(S \setminus T) \leqslant \sum_{i=1}^{\infty} \mu_\infty(S_i \setminus T) = 0.$$

Thus $\mu_\infty(S) = \mu_\infty(T) = \mu(G)$ where S is open in M and $G \subseteq S \times M'$. So we have shown the existence of a countable product M and open sets T and S and M such that

$$T \times M' \subseteq G \subseteq S \times M'; \mu(G) = \mu(T \times M') = \mu_\infty(T) = \mu(S \times M') = \mu_\infty(S).$$

Taking complements, if $F \subseteq X$ is closed, there is a countable product M and closed sets U, V in M such that

$$U \times M' \subseteq F \subseteq V \times M'; \mu(F) = \mu(U \times M') = \mu_\infty(U) = \mu(V \times M') = \mu_\infty(V).$$

Conclusion. By the above, and the regularity of μ, if $P \subseteq X$ is Borel,

$$\mu(P) = \inf_{P \subseteq S \times M'} \{\mu_\infty(S) | \text{open } S \subseteq M\} = \sup_{U \times M' \subseteq P} \{\mu_\infty(U) | \text{closed } U \subseteq M\}.$$

If then Q were Borel, the second equation shows that $\mu(Q) = 0$ (any U must be empty). In the first equation suppose $M = \prod_{j=1}^{\infty} I_{k_j}$ and $Q \subseteq S \times M'$. Then $\mu_\infty(\prod_{j=1}^{\infty} U_{k_j}) = 1$ so that $\mu_\infty(S) = 1$ and thus $\mu(Q) = 1$. This contradiction shows that Q cannot be Borel.

By starting from different Borel measures, other than Lebesgue measure, on $[0, 1]$, it is possible to prove by the same kind of method that other sets in X are not Borel. For example if $U_k = (a_k, b_k], 0 < a_k < b_k \leqslant 1$, then $\prod_k U_k$ is not Borel.

An Exponential Sum

6344 [1981, 352]. *Proposed by M. Machover, St. John's University.*

Let $\theta_1, \ldots, \theta_n$ be distinct angles (mod 2π). Let $A_1, \ldots, A_n$ be nonzero complex numbers. Let g be a function defined on the open interval (a, b) such that $(x, g(x))$ describes a continuous simple arc as x ranges over (a, b). Consider the sum

$$S(x) = \sum_{l}^{n} A_j \exp\big(i\big((\sin\theta_j)x + (\cos\theta_j)g(x)\big)\big).$$

(a) For each even integer $n > 2$ find A's, θ's, g, and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

(b) Show that, if $n = 3$, then there is no choice of A's, θ's, g and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

(c)* Conjecture: if n is odd, then there is no choice of A's, θ's, g and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

Solution by the proposer. (a) It is easy to see that $S(x) = 0$ for all $x \in (a, b)$ if $g(x) = 0$ for all $x \in (a, b)$ and

$$\theta_j = \pi - \theta_{n-j}, \quad A_j = -A_{n-j} \quad \text{for } j = 1, 2, \ldots, n/2.$$

(b) Suppose that $A_1 A_2 A_3 \neq 0$ and

$$\sum_{j=1}^{3} A_j \exp\big(i(x\sin\theta_j + y\cos\theta_j)\big) = 0 \tag{1}$$

where T is an open set in some countably infinite product $M = \prod_{j=1}^{\infty} I_{k_j}$ and $M' = \prod_{k \neq k_j} I_k$. Further $\mu(G) = \mu(T \times M') = \mu_\infty(T)$.

If now $G = \cup_\alpha J_\alpha$, $J_\alpha \in B$, let S_α be the projection of J_α into M so $S = \cup_\alpha S_\alpha$ will be the projection of G into M. S and the S_α are all open in M so that as M has a countable base for its open sets, there is a countable collection $S_1, S_2, \ldots$ of the S_α such that $S = \cup_{i=1}^{\infty} S_i$. Each $\mu_\infty(S_i \setminus T) = 0$ else we would have $\mu(G) > \mu(T \times M')$ and so

$$\mu_\infty(S \setminus T) \leqslant \sum_{i=1}^{\infty} \mu_\infty(S_i \setminus T) = 0.$$

Thus $\mu_\infty(S) = \mu_\infty(T) = \mu(G)$ where S is open in M and $G \subseteq S \times M'$. So we have shown the existence of a countable product M and open sets T and S and M such that

$$T \times M' \subseteq G \subseteq S \times M';\ \mu(G) = \mu(T \times M') = \mu_\infty(T) = \mu(S \times M') = \mu_\infty(S).$$

Taking complements, if $F \subseteq X$ is closed, there is a countable product M and closed sets U, V in M such that

$$U \times M' \subseteq F \subseteq V \times M';\ \mu(F) = \mu(U \times M') = \mu_\infty(U) = \mu(V \times M') = \mu_\infty(V).$$

Conclusion. By the above, and the regularity of μ, if $P \subseteq X$ is Borel,

$$\mu(P) = \inf_{P \subseteq S \times M'} \{\mu_\infty(S) | \text{open } S \subseteq M\} = \sup_{U \times M' \subseteq P} \{\mu_\infty(U) | \text{closed } U \subseteq M\}.$$

If then Q were Borel, the second equation shows that $\mu(Q) = 0$ (any U must be empty). In the first equation suppose $M = \prod_{j=1}^{\infty} I_{k_j}$ and $Q \subseteq S \times M'$. Then $\mu_\infty(\prod_{j=1}^{\infty} U_{k_j}) = 1$ so that $\mu_\infty(S) = 1$ and thus $\mu(Q) = 1$. This contradiction shows that Q cannot be Borel.

By starting from different Borel measures, other than Lebesgue measure, on [0, 1], it is possible to prove by the same kind of method that other sets in X are not Borel. For example if $U_k = (a_k, b_k], 0 < a_k < b_k \leqslant 1$, then $\prod_k U_k$ is not Borel.

An Exponential Sum

6344 [1981, 352]. *Proposed by M. Machover, St. John's University.*

Let $\theta_1, \ldots, \theta_n$ be distinct angles (mod 2π). Let $A_1, \ldots, A_n$ be nonzero complex numbers. Let g be a function defined on the open interval (a, b) such that $(x, g(x))$ describes a continuous simple arc as x ranges over (a, b). Consider the sum

$$S(x) = \sum_{l}^{n} A_j \exp\big(i\big((\sin\theta_j)x + (\cos\theta_j)g(x)\big)\big).$$

(a) For each even integer $n > 2$ find A's, θ's, g, and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

(b) Show that, if $n = 3$, then there is no choice of A's, θ's, g and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

(c)* Conjecture: if n is odd, then there is no choice of A's, θ's, g and (a, b) such that $S(x) = 0$ for all $x \in (a, b)$.

Solution by the proposer. (a) It is easy to see that $S(x) = 0$ for all $x \in (a, b)$ if $g(x) = 0$ for all $x \in (a, b)$ and

$$\theta_j = \pi - \theta_{n-j},\quad A_j = -A_{n-j}\quad \text{for } j = 1, 2, \ldots, n/2.$$

(b) Suppose that $A_1 A_2 A_3 \neq 0$ and

$$\sum_{j=1}^{3} A_j \exp\big(i(x\sin\theta_j + y\cos\theta_j)\big) = 0 \tag{1}$$

for $y = g(x)$, $a < x < b$. It suffices to show that two of the angles must be equal (mod 2π). We may rewrite (1) in the form

$$(2) \qquad A_1 \exp(i(x(\sin\theta_1 - \sin\theta_3) + y(\cos\theta_1 - \cos\theta_3))) \\ + A_2 \exp(i(x(\sin\theta_2 - \sin\theta_3) + y(\cos\theta_2 - \cos\theta_3))) = -A_3.$$

If we multiply each side of (2) by its conjugate and simplify we obtain

$$(3) \qquad x(\sin\theta_1 - \sin\theta_2) + y(\cos\theta_1 - \cos\theta_2) = \cos^{-1}a - \arg A_1 + \arg A_2$$

where $a = \dfrac{1}{2|A_1||A_2|}(|A_3|^2 - |A_1|^2 - |A_2|^2)$. Similarly we can deduce from (1) that

$$(4) \qquad x(\sin\theta_1 - \sin\theta_3) + y(\cos\theta_1 - \cos\theta_3) = \cos^{-1}b - \arg A_1 + \arg A_3$$

where $b = \dfrac{1}{2|A_1||A_3|}(|A_2|^2 - |A_1|^2 - |A_3|^2)$. The continuity of g ensures that the multiple-valued right-hand sides of (3) and (4) remain constant as x ranges over (a, b). Since the linear system (3), (4) is satisfied by every point (x, y) on the arc, the determinant of the system

$$(5) \qquad (\sin\theta_1 - \sin\theta_2)(\cos\theta_1 - \cos\theta_3) - (\sin\theta_1 - \sin\theta_3)(\cos\theta_1 - \cos\theta_2) \\ = \sin(\theta_3 - \theta_1) + \sin(\theta_2 - \theta_3) + \sin(\theta_1 - \theta_2) = 0.$$

It is readily seen that (5) implies that at least two of the angles are equal (mod 2π).

The proposer appeals to continuity to justify (5). His argument, however, remains valid for any g since the points (x, y) satisfying (3) and (4) form an uncountable set whereas the set of possible pairs of right-hand side values is countable. Thus there are at least two pairs of values (x, y) satisfying (3) and (4) with the same right-hand pair. Conjecture $(c)^*$ remains open.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Studies in Algebraic Geometry (*M.A.A. Studies in Mathematics, Vol.* 20). Edited by A. Seidenberg. Mathematical Association of America, Washington, D. C., 1980. xii + 143 pp. $16.00.

KENNETH R. MOUNT
Department of Mathematics, Northwestern University, Evanston, IL 60201

Tracing genealogy (teacher to student) is usually of questionable utility. In the case of the book edited by Seidenberg (O. Zariski's student), contemplating the line from the founders of algebraic geometry, that is from the founders as anointed by van der Waerden (one of the contributors to the volume), to the editor and contributors is at least amusing. For this family tree, the word algebraic in the phrase algebraic geometry is paramount.

The fundamental source of problems and techniques in algebraic geometry is the study of plane algebraic curves. A plane algebraic curve is the set of zeros in a plane (over the complex numbers) of an irreducible polynomial in two variables, or it is the projective version of such a curve which is the set of zeros in the projective plane (over the complex numbers) of a homogeneous polynomial in three variables.

Such studies have ancient roots, but most commentators on the history of algebraic geometry

for $y = g(x)$, $a < x < b$. It suffices to show that two of the angles must be equal (mod 2π). We may rewrite (1) in the form

$$(2) \qquad A_1 \exp(i(x(\sin\theta_1 - \sin\theta_3) + y(\cos\theta_1 - \cos\theta_3))) + A_2 \exp(i(x(\sin\theta_2 - \sin\theta_3) + y(\cos\theta_2 - \cos\theta_3))) = -A_3.$$

If we multiply each side of (2) by its conjugate and simplify we obtain

$$(3) \qquad x(\sin\theta_1 - \sin\theta_2) + y(\cos\theta_1 - \cos\theta_2) = \cos^{-1}a - \arg A_1 + \arg A_2$$

where $a = \dfrac{1}{2|A_1||A_2|}(|A_3|^2 - |A_1|^2 - |A_2|^2)$. Similarly we can deduce from (1) that

$$(4) \qquad x(\sin\theta_1 - \sin\theta_3) + y(\cos\theta_1 - \cos\theta_3) = \cos^{-1}b - \arg A_1 + \arg A_3$$

where $b = \dfrac{1}{2|A_1||A_3|}(|A_2|^2 - |A_1|^2 - |A_3|^2)$. The continuity of g ensures that the multiple-valued right-hand sides of (3) and (4) remain constant as x ranges over (a, b). Since the linear system (3), (4) is satisfied by every point (x, y) on the arc, the determinant of the system

$$(5) \qquad (\sin\theta_1 - \sin\theta_2)(\cos\theta_1 - \cos\theta_3) - (\sin\theta_1 - \sin\theta_3)(\cos\theta_1 - \cos\theta_2) = \sin(\theta_3 - \theta_1) + \sin(\theta_2 - \theta_3) + \sin(\theta_1 - \theta_2) = 0.$$

It is readily seen that (5) implies that at least two of the angles are equal (mod 2π).

The proposer appeals to continuity to justify (5). His argument, however, remains valid for any g since the points (x, y) satisfying (3) and (4) form an uncountable set whereas the set of possible pairs of right-hand side values is countable. Thus there are at least two pairs of values (x, y) satisfying (3) and (4) with the same right-hand pair. Conjecture $(c)^*$ remains open.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Studies in Algebraic Geometry (*M.A.A. Studies in Mathematics*, *Vol.* 20). Edited by A. Seidenberg. Mathematical Association of America, Washington, D. C., 1980. xii + 143 pp. $16.00.

KENNETH R. MOUNT
Department of Mathematics, Northwestern University, Evanston, IL 60201

Tracing genealogy (teacher to student) is usually of questionable utility. In the case of the book edited by Seidenberg (O. Zariski's student), contemplating the line from the founders of algebraic geometry, that is from the founders as anointed by van der Waerden (one of the contributors to the volume), to the editor and contributors is at least amusing. For this family tree, the word algebraic in the phrase algebraic geometry is paramount.

The fundamental source of problems and techniques in algebraic geometry is the study of plane algebraic curves. A plane algebraic curve is the set of zeros in a plane (over the complex numbers) of an irreducible polynomial in two variables, or it is the projective version of such a curve which is the set of zeros in the projective plane (over the complex numbers) of a homogeneous polynomial in three variables.

Such studies have ancient roots, but most commentators on the history of algebraic geometry

agree that one must consider the major development of the field to have started with the investigation of curves with more complex structure that that of the conics.

Shafarevich (c.f. [**2**]) points to questions asked and answered about integral calculus by the Bernoullis, Euler, and Abel as the first steps on the road to modern algebraic geometry. Euler's papers on elliptic integrals were written in the late 18th century and treated the question of integrating expressions of the form $dx/\sqrt{p(x)}$ when $p(x)$ is a polynomial of degree 4 or less. Specifically he asked for the solution of the differential equation

$$dx/\sqrt{p(x)} + dy/\sqrt{p(y)} = 0$$

and succeeded in finding the general solution. Euler catalogued the solutions for all the possible forms for $p(x)$; however, it is a bit easier to see why one looks upon this as a starting point for algebraic geometry if $p(x)$ is restricted to be a polynomial of degree 3 without multiple roots. One such example is $p(x) = 1 - x^3$. In the case of this example it is easy to see that the integral $\int dx/\sqrt{p(x)}$ can be easily interpreted as the integral $\int dx/y$ on the curve with equation

$$f(x, y) = y^2 - (1 - x^3) = 0.$$

This curve is topologically a torus with some holes at infinity. Normally one plugs the holes by embedding the curve in the projective plane and forming the closure of the embedded curve. The equation for this closure is relatively easy to find. For the example $f(x, y) = 0$, replace x by X/Z, replace y by Y/Z and multiply the result by Z^3. The result is the homogeneous equation

$$F = ZY^2 - (Z^3 - X^3) = 0.$$

Actually in this example the homogeneous equation is that of a nonsingular curve in the projective plane. The curve is nonsingular because the partials $\partial F/\partial X$, $\partial F/\partial Y$ and $\partial F/\partial Z$ do not simultaneously vanish at any point on the curve in projective space. For the curve with equation $f(x, y) = 0$ the differential dx/y determines a differential everywhere nonsingular on the projective curve. This means that for each point on the curve, together with a parameterization of the curve around the point, the differential can be written in the form $q\,dt$ where q is analytic at the point. In other words dx/y determines a differential of the first kind. A line integral of a differential of the first kind is an Abelian integral. No such differential of the first kind exists on the projective line, or equivalently, no such differentials of the first kind exist on nondegenerate conics. In fact the nonsingular cubic curves are topologically tori (genus 1 (real) surfaces) with a decidedly more intricate structure than the conics. Indeed the integral $\int dx/y$ cannot be written as a rational function in the simple functions x, $y \sin x$, $\log x$, etc. Euler did show that the sum

$$\int dx/\sqrt{p(x)} + \int dy/\sqrt{p(y)}$$

can be written as $\int dt/\sqrt{p(t)}$ and found an explicit representation for the function t.

In his studies, Abel cut the curve $y^2 = p(x)$ with a curve chosen from a linear family of curves and studied the sum of the integrals

$$\int_q^{q(1)} dx/y + \cdots + \int_q^{q(m)} dx/y$$

where the points $q(1), \ldots, q(m)$ are the points of intersection of the curve with a variable member of the family. The result of this investigation is one of the classic theorems of the subject from which one may derive, among other things, the general solution of the differential equation of Euler's studies. A quick classical discussion can be found in Lefschetz's book [**1**]. For Abel, the methods of study were those of the analyst.

In his 1857 paper on Abelian integrals, Riemann introduced the concept of the field of rational functions on a curve. For a projective curve this field consists of the restriction to the curve of all the functions with expressions P/Q where P and Q are homogeneous polynomials of the same

degree. Along with this concept he considered the idea of birational isomorphism. Curves are birationally isomorphic if the function fields are isomorphic (over the complex numbers). Conics are all birationally isomorphic to the projective line, for example.

In his discussion of the history of algebraic geometry (c.f. [3]) van der Waerden attributes the origin of algebraic geometry to M. Noether and the Italian school. The reason for this is that with these gentlemen algebraic techniques came to the fore. Noether's teacher Clebsch had begun the study of fields of rational functions on a surface. A surface is the set of zeros of a homogeneous polynomial in four variables. More formally stated, Clebsch began the study of the birational geometry of surfaces. Clebsch also had studied algebraic curves and had shown that the number of linearly independent Abelian integrals on a curve could be expressed in terms of the degree of the curve (i.e., the degree of the defining equation of the curve) and the number and type of singularities of the curve. This result expresses analytic (actually topological) invariants in terms of the algebraic (i.e., formal) properties of the curve.

It was M. Noether's position that Abelian integrals could be replaced with purely algebraic tools. He further asserted that it is the discussion of the field of functions of a curve or surface which is of principal interest. Noether used the concept of point group for his investigations of curves. A point group is a finite collection of points on a curve, each point with an assigned integer multiplicity. Modern geometers would call these point groups cycles while algebraists of the arithmetic school would call them divisors. As Abel had indicated the point groups of interest for the discussion of plane curves are those which are contained in the intersection of the curve with a second curve (with intersection multiplicities attached). For surfaces the natural generalization, used by Noether, was a discussion of the curves cut out on the surface by families of surfaces.

The ideas of Clebsch and Noether were developed and polished by the Italian school of geometry. Founders of this school were Cremona, C. Segre and Bertini. A student of Veronese and Segre, Castelnuovo, used numerical invariants to classify those surfaces with function fields isomorphic to the field of functions of the projective plane. Castelnuovo's student, Enriques, completed this birational classification of surfaces during the early part of this century.

The work of the Italian school, brilliant though it was, had some very shaky algebraic foundations. In particular the intersection theory, that is, the assignment of intersection multiplicities, was suspect. A major contributor to the work of shoring up the algebraic foundations of the Italian school was Emmy Noether, Max Noether's daughter. B. L. van der Waerden used Emmy Noether's work to build a solid algebraic basis for the theory of intersection. A short history of this development and the relations between van der Waerden's and E. Noether's work was published recently by van der Waerden in [3].

The contribution to Seidenberg's volume from van der Waerden is a discussion of the solution of one of the central issues of his intersection theory, using a powerful theorem proved by a student of Enriques, O. Zariski.

Much of the theory of Abelian integrals is best explained by noting that the nonsingular algebraic curves of genus 1 have a group structure. The multiplication on the curve can be described by rational functions which gives the curve the structure of an algebraic group. M. Rosenlicht, another student of O. Zariski, discusses some of the elementary theory of algebraic groups in his contribution to Seidenberg's collection. Abelian varieties, among which one finds the elliptic curves (curves of genus 1), are not discussed in this paper but Rosenlicht gives an elegant introduction to a second fundamental type of algebraic group, the linear algebraic group.

The development of modern intersection theory was also heavily influenced by A. Weil's work on the subject. Weil's proof of the Riemann hypothesis for finite function fields used the intersection theory and correspondence principles of the Italian school. The principles were used to study curves defined over fields of nonzero characteristic which meant that an intersection theory in nonzero characteristic was crucial. The multiplicity theory of van der Waerden was applicable in nonzero characteristic; however, Weil's development of intersection theory is distinct

from van der Waerden's and was firmly rooted in the theory of local rings and power series rings. It is Weil's theory, and the local geometry of Zariski, Chevalley, and their disciples, that motivated a great deal of the commutative algebra that has developed since 1950.

The study of the relation between local theory and the global results on multidimensional varieties has blossomed into the Serre-Grothendieck school wherein homological algebra, linear algebra run amuck, plays a key role. The article by J. Ohm, a student of Seidenberg, gives the reader a taste of this rich broth of commutative algebra and homological methods.

S. Kleiman, another of the Zariski-Harvard school, is the author of the last paper in the collection. Kleiman leads the reader into the subject of the enumerative theory of conics where again intersection theory plays a major role. In this paper one is carried from the geometry of the ancient Greeks to the modern discussions of intersection theory by Grothendieck.

It's a lovely collection—enjoy.

References

1. S. Lefschetz, Algebraic Geometry, Princeton Univ. Press, 1953.

2. I. R. Shafarevich, Basic Algebraic Geometry, Grundlehren der Mathematischen Wissenschaften Band 213, Springer-Verlag, New York, 1974.

3. B. L. van der Waerden, Foundations of Algebraic Geometry, Archive for History of Exact Sciences, vol. 7, no. 3, 1971, p. 171.

Conformal Mappings on Riemann Surfaces. By Harvey Cohn. Dover, New York, reprinted 1980. pp. ix + 325, $6.00.

JAMES A. JENKINS
Department of Mathematics, Washington University, St. Louis, MO 63130

In a book designed to provide an introduction to the theory of Riemann surfaces, the reviewer would expect to find an essential core of material consisting of the following items.

I. The definition of a Riemann surface as a complex analytic manifold of dimension one; that is, a connected Hausdorff space covered by domains D_i, each of which has an associated homeomorphic mapping ϕ_i onto a plane disc, so that for overlapping D_i and D_j, the map $\phi_j\phi_i^{-1}$ is a conformal mapping of $\phi_i\,(D_i \cap D_j)$ (i.e., a (1, 1) mapping by a regular (= holomorphic) function). The ϕ_i are called local uniformizing parameters, and the D_i are called coordinate neighborhoods.

II. The existence of meromorphic functions and differentials on a Riemann surface. A complex valued function f defined on a subdomain of a Riemann surface is called meromorphic if, for every local uniformizing parameter ϕ_i, $f\phi_i^{-1}$ is meromorphic. A meromorphic linear differential assigns to each coordinate neighborhood D_i a meromorphic function $f_i(z)$ in $\phi_i(D_i)$ with the property that if D_i and D_j are overlapping coordinate neighborhoods, then $f_i(z) = f_j(w)\frac{dw}{dx}$, where $w = \phi_j\phi_i^{-1}(z)$. More generally, a linear differential assigns to each D_i two functions u, v, which transform so that $udx + vdy (z = x + iy)$ is invariant. A second order differential is given by one function C which transforms so that $C\,dx\,dy$ is invariant. (The exterior product of two linear differentials gives a second order differential.) A harmonic function h on a subdomain of a Riemann surface is one for which for each local uniformizing parameter ϕ_i, $h\phi_i^{-1}$ is harmonic. A harmonic differential is locally the differential of a harmonic function. One can also speak of a regular mapping f of a Riemann surface R into a Riemann surface S, requiring that for appropriate local uniformizing parameters, $\phi_j f \phi_i^{-1}$ be regular.

III. Covering surfaces and uniformization. For Riemann surfaces R and S, R is called a

from van der Waerden's and was firmly rooted in the theory of local rings and power series rings. It is Weil's theory, and the local geometry of Zariski, Chevalley, and their disciples, that motivated a great deal of the commutative algebra that has developed since 1950.

The study of the relation between local theory and the global results on multidimensional varieties has blossomed into the Serre-Grothendieck school wherein homological algebra, linear algebra run amuck, plays a key role. The article by J. Ohm, a student of Seidenberg, gives the reader a taste of this rich broth of commutative algebra and homological methods.

S. Kleiman, another of the Zariski-Harvard school, is the author of the last paper in the collection. Kleiman leads the reader into the subject of the enumerative theory of conics where again intersection theory plays a major role. In this paper one is carried from the geometry of the ancient Greeks to the modern discussions of intersection theory by Grothendieck.

It's a lovely collection—enjoy.

References

1. S. Lefschetz, Algebraic Geometry, Princeton Univ. Press, 1953.

2. I. R. Shafarevich, Basic Algebraic Geometry, Grundlehren der Mathematischen Wissenschaften Band 213, Springer-Verlag, New York, 1974.

3. B. L. van der Waerden, Foundations of Algebraic Geometry, Archive for History of Exact Sciences, vol. 7, no. 3, 1971, p. 171.

Conformal Mappings on Riemann Surfaces. By Harvey Cohn. Dover, New York, reprinted 1980. pp. ix + 325, \$6.00.

JAMES A. JENKINS
Department of Mathematics, Washington University, St. Louis, MO 63130

In a book designed to provide an introduction to the theory of Riemann surfaces, the reviewer would expect to find an essential core of material consisting of the following items.

I. The definition of a Riemann surface as a complex analytic manifold of dimension one; that is, a connected Hausdorff space covered by domains D_i, each of which has an associated homeomorphic mapping ϕ_i onto a plane disc, so that for overlapping D_i and D_j, the map $\phi_j\phi_i^{-1}$ is a conformal mapping of $\phi_i\,(D_i \cap D_j)$ (i.e., a (1, 1) mapping by a regular (= holomorphic) function). The ϕ_i are called local uniformizing parameters, and the D_i are called coordinate neighborhoods.

II. The existence of meromorphic functions and differentials on a Riemann surface. A complex valued function f defined on a subdomain of a Riemann surface is called meromorphic if, for every local uniformizing parameter ϕ_i, $f\phi_i^{-1}$ is meromorphic. A meromorphic linear differential assigns to each coordinate neighborhood D_i a meromorphic function $f_i(z)$ in $\phi_i(D_i)$ with the property that if D_i and D_j are overlapping coordinate neighborhoods, then $f_i(z) = f_j(w)\frac{dw}{dx}$, where $w = \phi_j\phi_i^{-1}(z)$. More generally, a linear differential assigns to each D_i two functions u, v, which transform so that $udx + vdy(z = x + iy)$ is invariant. A second order differential is given by one function C which transforms so that $C\,dx\,dy$ is invariant. (The exterior product of two linear differentials gives a second order differential.) A harmonic function h on a subdomain of a Riemann surface is one for which for each local uniformizing parameter ϕ_i, $h\phi_i^{-1}$ is harmonic. A harmonic differential is locally the differential of a harmonic function. One can also speak of a regular mapping f of a Riemann surface R into a Riemann surface S, requiring that for appropriate local uniformizing parameters, $\phi_j f\phi_i^{-1}$ be regular.

III. Covering surfaces and uniformization. For Riemann surfaces R and S, R is called a

covering surface of S if there exists a regular mapping π of R into S. The surface R is called an unbranched covering if π is a local homeomorphism, an unbordered covering if for any path α in S with initial point A, and B in R covering A ($\pi B = A$), there is a path β in R with $\pi\beta = \alpha$. The covering surfaces of S have an obvious hierarchical structure. More particularly, for unbranched, unbordered coverings this corresponds to subgroup structure of the fundamental group. For these there is a maximal element which is simply-connected—the universal covering surface.

Uniformization of a Riemann surface R means finding a homeomorphic mapping into the sphere which provides a local uniformizing parameter for a neighborhood of each point. This is possible only if every simple closed curve on R separates R (such a surface is called "schlichtartig"). Koebe showed that in this case it is possible to prove uniformization by the study of a particular meromorphic differential. A simply connected Riemann surface is schlichtartig, however, so that the universal covering surface U of any Riemann surface R admits such a mapping. The image can be taken as the sphere, the plane or a disc, the first two occurring only in special cases. The covering transformations of U (self-conformal mappings ϕ with $\pi\phi = \pi$) correspond in the general case to linear transformations of the disc D. They form a group Γ such that identifying points of D congruent under Γ, we obtain a Riemann surface conformally equivalent to R.

IV. The Riemann-Roch Theorem. On any Riemann surface if we assign lower bounds for the orders of zeros and poles (counted as negative) for either meromorphic functions or meromorphic linear differentials at an isolated set of points, we obtain a complex linear space. On a compact Riemann surface each of these spaces has finite dimension. If we take bounds for the orders of zeros and poles of functions and of differentials satisfying a certain complementation condition, then the dimensions of the two spaces differ only by an amount depending on the sums of the bounds. This result has many important applications.

Additional contents of the book might take several forms. There are many directions in which one could provide extensions of these central results. On the other hand, space might be devoted to motivation (of the basic definitions and the entities to be developed) or the provision of prerequisite material.

The author states a very special purpose for the book in hand, namely to provide an introduction to Riemann surface theory for a student whose preparation consists of "an ordinary semester of complex analysis." Comparing the contents with those suggested above, we find that the existence proofs are given only for a finite triangulated Riemann surface (finite means either compact or equivalent to a subdomain of a compact surface bounded by a finite number of contours). Covering surfaces in the general sense are mentioned only very sketchily in an appendix, and there is no general statement concerning uniformization. Thus we have the interesting phenomenon of a book on Riemann surfaces with forty-five bibliographical references in which the name of Koebe does not appear. Not surprisingly there is no attempt to go beyond the basic material and evidently the author's big thing is motivation. It is never stated explicitly that the book is a fairly faithful presentation of the author's lectures but in the absence of contrary evidence it is reasonable to assume that. This being so, it seems a student would find the recitation of results in the first three chapters terribly boring. Particularly disappointing is the lacklustre treatment of analytic continuation since the best motivation for the basic concepts and definitions of the theory of Riemann surfaces is provided by a good treatment of this topic, for example like that in Ahlfors' Complex Analysis (preferably the first or second edition; the third edition seems in need of a good dose of antibiotics).

As for provision of prerequisite material, the author makes no organized attempt in this direction in fields other than complex variables. Indeed, with one exception, very little need be provided. Only the simplest concepts of point set topology and algebra are needed. (If one uses the method of orthogonal projection, a few basic results from Lebesgue theory are required. The author avoids this by using the alternating method from potential theory but with an attendant loss of generality.) The exception is the subject which the reviewer knows by the name of algebraic

topology. One can hardly assume a knowledge of this, particularly in the somewhat old-fashioned form in which it is most useful in the study of Riemann surfaces, even with students more mature than those envisioned by the author. The author does give a rather offhand introduction to homology but it is quite distressing to find the presumably novice students being referred in cavalier fashion to "textbooks in topology" and "the literature."

The book is liberally provided with exercises. This is by no means an unmixed blessing since the author is consistently guilty of writing the book in the exercises. In the instance of a student who has expert guidance, a case can be made in support of this procedure (though not so as to convince the reviewer), but for one attempting to master the book independently it can be terminally frustrating, despite the promotional blurb on the back. (Can publishers be arraigned for lack of truth in advertising?)

The book does contain certain nice insights, but they seem best suited to someone who has already a reasonable knowledge of the material (as opposed to the stated intended audience). On the other hand, there are entirely unnecessary misstatements (in distinction to "compromises").

If, then, it is desirable to learn of the theory of Riemann surfaces as early as possible, is there any alternative? The reviewer shares the author's belief in the importance of the subject though not necessarily his sense of urgency. A student would be much better off first to take a good basic one-year course in Function Theory. This would provide a full understanding of the material in the first part of this book. Then, in one semester, it would be possible to give a very presentable introduction to the theory of Riemann surfaces. If two semesters could be employed, it would be possible to give a thoroughgoing, mathematically complete treatment extending far beyond the material of this book.

For a person who wants only some idea of what Riemann surfaces are about, the book has far too much detail. Someone who really wishes to understand the mathematics of Riemann surfaces would not obtain the desired complete development and would get no clear idea of the nature of mathematical proof. In teaching a course on Riemann surfaces, to follow the author's format would be at best unwise, and possibly disastrous.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

A CUPM panel presented an excellent report on the topic of minimal mathematical competencies for college graduates (this MONTHLY, vol. 89, pp. 266-272).

Generally, I am in agreement with the report, but there is one modification that I wish to recommend, namely, to change topic 14, which currently is "statistics and its dangers." Lists published by CUPM panels tend to become the "bible" on a topic. In this case "verse" 14 should be changed to reflect a modern concept of statistics and promote a positive attitude about the subject. Just as mathematicians and accountants are not clearly separated by some (as indicated by the CUPM panel), so also are professional statisticians and sports "statisticians" confused. Moreover, the phrase "statistics and its dangers" indirectly propagates the misconception that one can prove anything with statistics. Obviously, statistics is no more dangerous than percent (topic 9) or rounding off (topic 5).

A danger in topic 14 now is its potential for misleading faculty who develop the proposed course. For example, an instructor might take 14 literally and inappropriately expend all effort

topology. One can hardly assume a knowledge of this, particularly in the somewhat old-fashioned form in which it is most useful in the study of Riemann surfaces, even with students more mature than those envisioned by the author. The author does give a rather offhand introduction to homology but it is quite distressing to find the presumably novice students being referred in cavalier fashion to "textbooks in topology" and "the literature."

The book is liberally provided with exercises. This is by no means an unmixed blessing since the author is consistently guilty of writing the book in the exercises. In the instance of a student who has expert guidance, a case can be made in support of this procedure (though not so as to convince the reviewer), but for one attempting to master the book independently it can be terminally frustrating, despite the promotional blurb on the back. (Can publishers be arraigned for lack of truth in advertising?)

The book does contain certain nice insights, but they seem best suited to someone who has already a reasonable knowledge of the material (as opposed to the stated intended audience). On the other hand, there are entirely unnecessary misstatements (in distinction to "compromises").

If, then, it is desirable to learn of the theory of Riemann surfaces as early as possible, is there any alternative? The reviewer shares the author's belief in the importance of the subject though not necessarily his sense of urgency. A student would be much better off first to take a good basic one-year course in Function Theory. This would provide a full understanding of the material in the first part of this book. Then, in one semester, it would be possible to give a very presentable introduction to the theory of Riemann surfaces. If two semesters could be employed, it would be possible to give a thoroughgoing, mathematically complete treatment extending far beyond the material of this book.

For a person who wants only some idea of what Riemann surfaces are about, the book has far too much detail. Someone who really wishes to understand the mathematics of Riemann surfaces would not obtain the desired complete development and would get no clear idea of the nature of mathematical proof. In teaching a course on Riemann surfaces, to follow the author's format would be at best unwise, and possibly disastrous.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

A CUPM panel presented an excellent report on the topic of minimal mathematical competencies for college graduates (this MONTHLY, vol. 89, pp. 266-272).

Generally, I am in agreement with the report, but there is one modification that I wish to recommend, namely, to change topic 14, which currently is "statistics and its dangers." Lists published by CUPM panels tend to become the "bible" on a topic. In this case "verse" 14 should be changed to reflect a modern concept of statistics and promote a positive attitude about the subject. Just as mathematicians and accountants are not clearly separated by some (as indicated by the CUPM panel), so also are professional statisticians and sports "statisticians" confused. Moreover, the phrase "statistics and its dangers" indirectly propagates the misconception that one can prove anything with statistics. Obviously, statistics is no more dangerous than percent (topic 9) or rounding off (topic 5).

A danger in topic 14 now is its potential for misleading faculty who develop the proposed course. For example, an instructor might take 14 literally and inappropriately expend all effort

showing cute examples of how to mislead others with the misuse of statistics. The CUPM panel's recommendation that this course be taught by an experienced senior faculty member is very important. Furthermore, I recommend that this person have experience in teaching statistics.

Taking a broader view, one sees that those at the forefront of statistics are beginning to discuss "Statistics as a Discipline." In fact, this was the title of the main address presented by Paul Minton at the second annual Conference for Texas Statisticians held at the Baylor University (April, 1982). The main point of Professor Minton's talk concerned the idea of offering *undergraduate degrees* in statistics.

There are two points I want to make. First, it is conceivable that a statistics course "for coping with life" could cover most of the topics in the list given by the CUPM panel. Second, topic 14 in a list of topics for this statistics course could be "mathematics and its dangers." Need I say more?

Since it would be cowardly to avoid a recommendation for a replacement for topic 14, I suggest:

14. The Use and Interpretation of Statistics in Everyday Life.

Danny W. Turner
Department of Mathematics
Baylor University
Waco, Texas 76798

MISCELLANEA

With Apologies to John Milton

94. John Milton wrote a famous sonnet on being blind. This modern version looks at a different kind of blindness.

When I consider how my grants are spent,
Ere half my days in these dark halls and wide;
My teaching talent, which I'd love to hide,
Lodged with me useless, though my soul more bent
To serve therewith my Provost, and present
My teaching ratings, lest he returning chide;
"Do Deans exact much teaching, grants denied?"
I fondly ask: the Provost, to prevent
That murmur, soon replies, "Deans do not need
Either that teaching or those grants. Who best
Does his committee work, serves best. His state
Is Deanly: thousands at his bidding speed
And fly to Washington, disdaining rest;
They also serve who only loaf and wait."

—Edwin Hewitt

ANSWER TO "PHOTO" ON PAGE 100

The subject is Max Zorn; the picture was taken in the summer of 1981. He is alive and well and very much a part of the mathematical community in Bloomington, Indiana.

showing cute examples of how to mislead others with the misuse of statistics. The CUPM panel's recommendation that this course be taught by an experienced senior faculty member is very important. Furthermore, I recommend that this person have experience in teaching statistics.

Taking a broader view, one sees that those at the forefront of statistics are beginning to discuss "Statistics as a Discipline." In fact, this was the title of the main address presented by Paul Minton at the second annual Conference for Texas Statisticians held at the Baylor University (April, 1982). The main point of Professor Minton's talk concerned the idea of offering *undergraduate degrees* in statistics.

There are two points I want to make. First, it is conceivable that a statistics course "for coping with life" could cover most of the topics in the list given by the CUPM panel. Second, topic 14 in a list of topics for this statistics course could be "mathematics and its dangers." Need I say more?

Since it would be cowardly to avoid a recommendation for a replacement for topic 14, I suggest:

14. The Use and Interpretation of Statistics in Everyday Life.

Danny W. Turner
Department of Mathematics
Baylor University
Waco, Texas 76798

MISCELLANEA

With Apologies to John Milton

94. John Milton wrote a famous sonnet on being blind. This modern version looks at a different kind of blindness.

When I consider how my grants are spent,
Ere half my days in these dark halls and wide;
My teaching talent, which I'd love to hide,
Lodged with me useless, though my soul more bent
To serve therewith my Provost, and present
My teaching ratings, lest he returning chide;
"Do Deans exact much teaching, grants denied?"
I fondly ask: the Provost, to prevent
That murmur, soon replies, "Deans do not need
Either that teaching or those grants. Who best
Does his committee work, serves best. His state
Is Deanly: thousands at his bidding speed
And fly to Washington, disdaining rest;
They also serve who only loaf and wait."

—Edwin Hewitt

ANSWER TO "PHOTO" ON PAGE 100

The subject is Max Zorn; the picture was taken in the summer of 1981. He is alive and well and very much a part of the mathematical community in Bloomington, Indiana.

Developmental Delivery Systems That Deliver

A diagnostic/prescriptive approach . . .

PCSP: Personalized Computational Skills Program
A Brush-Up Book®

Bryce R. Shaw, Senior Author
Gwen Shelton, Bee County College, Texas
Sandra Pryor Clarkson, Hunter College, New York

VOLUME 1 COMPUTATIONAL SKILLS • 1980
Module A: Whole Numbers
Module B: Fractions
Module C: Decimals, Proportions, and Percents

VOLUME 2 APPLICATIONS • 1981
Module D: Applications Using Computational Skills
Module E: Special Skills with Applications
Module F: Beginning Algebra with Applications

PCSP COMBINED EDITION • 1982

PCSP uses skill models, worked examples, and thousands of practice problems and applications to improve computational competency. Built-in diagnostic and mastery tests pinpoint students' problem areas and measure their learning progress.

The six modules are available separately or in two volumes, which may be used independently, concurrently, or in sequence. The Combined Edition offers a quick brush-up on basic skills together with their applications in a single volume.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 3 March 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

ALGEBRA, GEOMETRY, AND ALGEBRAIC GEOMETRY: SOME INTERCONNECTIONS

KEITH M. KENDIG

Department of Mathematics, Cleveland State University, Cleveland, OH 44115

1. Introduction. For centuries, men have been intrigued by the interplay between algebra and geometry; many important advances in mathematics have essentially come down to linking up the two in one way or other. The ancient Greeks established such a link when they found straightedge-and-compass constructions for the sum, difference, product, quotient and square root of lengths. Closer to our own time, Poincaré did it when he associated groups (algebra) with topological manifolds (geometry); of course this has grown into the whole area of algebraic topology.

Possibly the single greatest step in connecting up algebra and geometry was Descartes introduction in 1637 of Cartesian geometry (or "analytic geometry," as Lacroix named it in 1792). It laid the mathematical foundation for the calculus and Newtonian physics a half century later. It put algebra and geometry on a more equal footing, and because it set up a translation between algebra and geometry, it gave our imagination "two ends"—an algebraic one and a geometric one; geometric insight could often be translated into an algebraic one, and vice versa. Still today, an important part of a student's exposure to analytic geometry consists in making the translations of algebra-to-geometry ("sketch $2x^2 + xy + 3y^2 = 0$"), and geometry-to-algebra ("find an equation of the set of points equidistant from the y-axis and the point $(2,3)$").

Powerful though Descartes' union of algebra and geometry is, it is only the beginning of a much larger story. In most textbooks today, analytic geometry is done in the real setting of $\mathbb{R}^2$ or $\mathbb{R}^3$. But one may replace these by other objects (complex n-space, $\mathbb{C}^n$, is just one example), and try to do analytic geometry in this new context. The extremely fundamental nature of this "algebra-geometry" linkage asserts itself, and unexpected connections, inspiration and enlightenment come pouring out. In a sense, exploring the cartesian-geometry type of connection between geometry and algebra (especially at the polynomial level) is "algebraic geometry."

2. A New Setting for Analytic Geometry. Analytic geometry was born, and grew, over the reals. But such a context has inherent in it certain limitations; in particular, it fails to give us "closure," in both an algebraic and a geometric sense. (We explain in a moment.) Each of these can lead to "exceptions." Since important parts of the history of mathematics can be regarded as the process of removing various kinds of exceptions, it is interesting to extend space to one that is closed algebraically and geometrically, and then do analytic geometry in it. This leads to "classical algebraic geometry." We explain.

First, $\mathbb{R}$ fails to be closed algebraically. Thus, for example, the real polynomial $x^2 - a$ has a zero in $\mathbb{R}$ (visually, the parabola $y = x^2 - a$ intersects the x-axis $\mathbb{R}_x$) *except* when a is negative. Extending $\mathbb{R}$ to its algebraic completion $\mathbb{C}$ removes this exceptional behavior, and now any polynomial (even with coefficients in $\mathbb{C}$) has a zero (i.e., the graph in $\mathbb{C}_{xy}$ of $y = p(x)$ intersects the x-axis $\mathbb{C}_x$)—*except* when the polynomial is a nonzero constant. (This is the Fundamental Theorem of Algebra.) This last exception arises because the two parallel copies of $\mathbb{C}$ don't meet. But this kind of exception can be removed also, by completing geometrically, adding a "point at infinity" to each complex line $\mathbb{C}$ in $\mathbb{C}^2$. For this, let $ax + by = 0$ $(a, b \in \mathbb{C})$ be any complex 1-space of $\mathbb{C}^2$. We add an "ideal" point P_∞ to it; we may think of this extension of $\mathbb{C}$ as a "Riemann sphere." Now add this *same* point P_∞ to each of the complex lines $ax + by = c$

Keith Kendig is an algebraic geometer with a side interest in computers. He received his Ph.D. from UCLA and spent two years at the Institute for Advanced Study in Princeton; he is the author of *Elementary Algebraic Geometry* (Springer-Verlag, Graduate Texts in Mathematics vol. 44). He also coauthored a book with his wife Joan on vegetarian cooking: each recipe in it is computer-tested for protein balance and nutritional soundness, and the couple appeared on a number of TV shows around the U.S. He is a regular jogger and a musician. He plays the cello in a string quartet which has given various performances in the Cleveland area. His current interest is in using computers as an aid to composers in writing music.

($c \in \mathbb{C}$) (i.e., to all the lines parallel to $ax + by = 0$). Since P_∞ is the only point common to these lines, it is the point in which any two meet. Adding one such point to (all lines parallel to) *each* 1-space $ax + by = 0$ ($a, b \in \mathbb{C}$) then extends $\mathbb{C}^2$ to what is called the *complex projective plane* $\mathbb{P}^2(\mathbb{C})$; it is manufactured so that any two distinct completed lines in it meet in exactly one point. It turns out that all the added points at infinity themselves form a complex line called "the line at infinity"; this line intersects any other line in one point. One can now give $\mathbb{P}^2(\mathbb{C})$ a topology by first observing that the points (respectively lines) of $\mathbb{P}^2(\mathbb{C})$ may be identified with the 1-subspaces (respectively 2-subspaces) of $\mathbb{C}^3$. (Thus "any two distinct lines of $\mathbb{P}^2(\mathbb{C})$ intersect in a point" translates to "any two distinct 2-subspaces of $\mathbb{C}^3$ intersect in a 1-subspace.") One may now identify "open neighborhood in $\mathbb{P}^2(\mathbb{C})$" with "the set of 1-subspaces in $\mathbb{C}^3$ intersecting an open set in $\mathbb{C}^3$". This topology on $\mathbb{P}^2(\mathbb{C})$ also induces one on the completed lines; they are in fact topological spheres.

Now to do analytic geometry in $\mathbb{P}^2(\mathbb{C})$, we surely want to be able to sketch curves there, just as we did in $\mathbb{R}^2$. The idea is simple: first, any nonconstant polynomial $p(x, y)$ in $\mathbb{C}[x, y]$ defines an *algebraic curve* (or simply a *curve*) C in $\mathbb{C}_{xy}$, consisting of the set of all points in $\mathbb{C}_{xy}$ at which $p(x, y)$ is zero. Then the topological closure of C in $\mathbb{P}^2(\mathbb{C})$ is called *the (algebraic) curve in* $\mathbb{P}^2(\mathbb{C})$ *defined by* $p(x, y)$. At this new level, sketching some simple curves will throw some unexpected light on their real analogues.

A little later, we will see how extending $\mathbb{C}^2$ to $\mathbb{P}^2(\mathbb{C})$ not only removes an exception from the statement of the Fundamental Theorem of Algebra, but also leads to a symmetric and much more general form of the theorem.

3. A Basic Example. In this section we sketch the curve in $\mathbb{P}^2(\mathbb{C})$ defined by $y - x^2$. We first sketch the "complex parabola" $y = x^2$ in $\mathbb{C}_{xy}$; then we take its topological closure in $\mathbb{P}^2(\mathbb{C})$. We begin by noting that the graph in $\mathbb{C}_{xy}$ of $y = x^2$ is homeomorphic to $\mathbb{C}_x$. Therefore the graph is a real two-dimensional surface in $\mathbb{C}_{xy} = \mathbb{R}^4$. To find out what this real surface looks like, we write $x = x_1 + ix_2$, $y = y_1 + iy_2$, and $\mathbb{C}^2 = \mathbb{R}_{x_1x_2x_3x_4}$. Then $y = x^2$ becomes $y_1 + iy_2 = (x_1 + ix_2)^2$. Equating real and imaginary parts gives us

$$(*)\qquad \begin{cases} y_1 = x_1^2 - x_2^2 \\ y_2 = 2x_1x_2 \end{cases}$$

Our surface, then, is the set of points in $\mathbb{R}_{x_1x_2y_1y_2}$ simultaneously satisfying the equations in $(*)$.

Because this surface lies in $\mathbb{R}^4$, it is hard to visualize directly. Now an inhabitant of "Flatland" might visualize an object in $\mathbb{R}^3$ by letting one coordinate be "time," and piecing together the various slices corresponding to time = a constant. We can do the same, letting x_2, for example, play the role of time and looking at the part of the surface in each three-dimensional slice. Each such part is a real curve; we will then fit the curves together to get the surface.

We consider first $x_2 = 0$. The equation $y_2 = 2x_1x_2$ of $(*)$ becomes $y_2 = 0$, so the curve lies in the plane $\mathbb{R}_{x_1y_1}$. The other equation in $(*)$ becomes $y_1 = x_1^2$; therefore the curve for $x_2 = 0$ is just the original real parabola in $\mathbb{R}_{x_1y_1}$. When $x_2 = a \neq 0$, the equations $(*)$ again define a real parabola in the corresponding 3-dimensional slice; it lies in the plane $y_2 = 2ax_1$. Now notice that our basic equation $y = x^2$ implies $|y|/|x| \to \infty$ as $|x| \to \infty$. From this one can easily deduce that for each $x_2 \in \mathbb{R}$, the corresponding real parabolas all have this property in common: each branch approaches the point at infinity added to $\mathbb{C}_y$. (Let us call it P_∞.) After taking the topological closure, we see that for each $x_2 = a \in \mathbb{C}$, the corresponding parabola together with P_∞ forms a loop; any two such loops touch at the one point P_∞. We may topologically redraw all this as in Fig. 1 where, say, the heavily-drawn curve corresponds to the original real parabola, the lightly-drawn curves correspond to $x_2 = a > 0$, and the dotted curves on the other side of the surface, to $x_2 = a < 0$.

Thus the variety in $\mathbb{P}^2(\mathbb{C})$ defined by $y^2 = x$ is topologically a *sphere*. In the ordinary real part, we only see a very small part of this: one loop, minus P_∞.

4. Some More Examples. We may apply the above method to sketch other graphs. Let's look briefly at the complex circle defined in $\mathbb{C}_{xy}$ by $x^2 + y^2 = 1$. This equation yields these two real

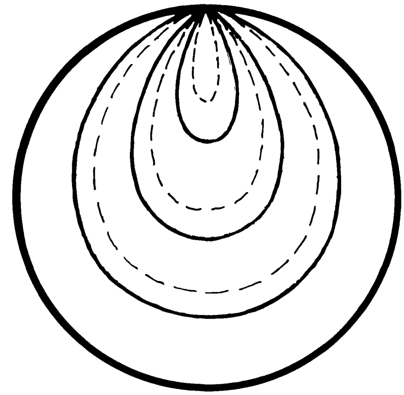


FIG. 1

equations:

$$(**)\qquad \begin{cases} x_1^2 - x_2^2 + y_1^2 - y_2^2 = 1 \\ 2x_1x_2 + 2y_1y_2 = 0 \end{cases}$$

Consider first $x_2 = 0$. Then $2x_1x_2 + 2y_1y_2 = 0$ becomes $y_1y_2 = 0$, which implies that $y_1 = 0$ or $y_2 = 0$. From this it follows that the part of our surface in the 3-dimensional slice $x_2 = 0$ consists of the union of the real circle $x_1^2 + y_1^2 = 1$ in $\mathbb{R}_{x_1y_1}$ and the hyperbola $x_1^2 - y_2^2 = 1$ in $\mathbb{R}_{x_1y_2}$. The branches of the hyperbola intersect the line at infinity in two points (the two points in which its asymptotes intersect the line at infinity). Adding these two points makes the hyperbola into a topological circle, so we now have two topological circles touching at two points. We let the circle $x_1^2 + y_1^2 = 1$ be the equator of a sphere, and the completed hyperbola, a great circle passing through the north and south poles. These two great circles divide the sphere into four quarters. One may show that the curves corresponding to $x_2 = a > 0$ fill in two opposite quarters; the curves corresponding to $x_2 = a < 0$ fill in the other two quarters. We see from this that the real part—what we see in ordinary analytic geometry—is just the tip of an iceberg; working in $\mathbb{P}^2(\mathbb{C})$ reveals the rest of it to us.

This is a very powerful approach, because often the real part hardly reveals even the tip of the iceberg! For example $x^2 + y^2 = 0$, a "circle of radius 0," is just a single point in $\mathbb{R}^2$. Looked at in $\mathbb{P}^2(C)$, we see that the real circle $x_1^2 + y_1^2 = 1$ of the previous example has now shrunk to a point, and the hyperbola $x_1^2 - y_2^2 = 1$ has reduced to its asymptotes $x_1^2 - y_2^2 = 0$. It appears that the equator of our sphere has been squeezed to a point, leaving two spheres touching at that point. Is this actually the topological picture of a "circle of radius 0"? We may look at it this way: Since $x^2 + y^2$ equals $(x + iy)(x - iy)$, and since this product is zero iff one of the factors is, we see that our 0-radius circle is the union of two complex lines intersecting at the origin. After adding the point at infinity to each such line, we get the union of two spheres touching at one point. Notice that as soon as r becomes 0, $x^2 + y^2 - r^2$ becomes reducible over $\mathbb{C}$ (one can verify that $x^2 + y^2 - r^2$ is irreducible whenever $r \neq 0$); and, in beautiful consonance with this, the associated topological object becomes in a sense "reducible" (two touching spheres) instead of a single "irreducible" sphere. This whole idea is revealed to us when we work in $\mathbb{P}^2(\mathbb{C})$. It was kept secret in $\mathbb{R}^2$.

But we need not stop here! How about a "circle of imaginary radius," say $x^2 + y^2 = -1$? In $\mathbb{R}^2$ we get the empty set. However in $\mathbb{P}^2(\mathbb{C})$ we once again get a full picture: $x^2 + y^2 + 1$ is irreducible, and the "sketch" turns out again to be topologically a sphere. In fact, *any* quadratic irreducible over $\mathbb{C}$ gives us a sphere, and any quadratic reducing to two distinct factors gives us two touching spheres.

5. Donuts and Coffee Cups. The reader might well wonder if *every* curve in $\mathbb{P}^2(\mathbb{C})$ defined by a polynomial consists of a sphere or spheres. Or can we get donuts, two-handled spheres, and even surfaces of higher genus?

Let us consider the exercise, often encountered in ordinary analytic geometry, of graphing $y^2 = x(x^2 - 1)$. In $\mathbb{R}^2$, the graph is "symmetric about the x-axis," and there are two "excluded regions" lying above the two open intervals $(0,1)$ and $(-\infty, -1)$. This real part appears in the x_1y_1-plane of Fig. 2. At time $x_2 = 0$ there's also a "mirror image" of that curve in the x_1y_2-plane. (Those exceptional "excluded regions" really aren't excluded after all.)

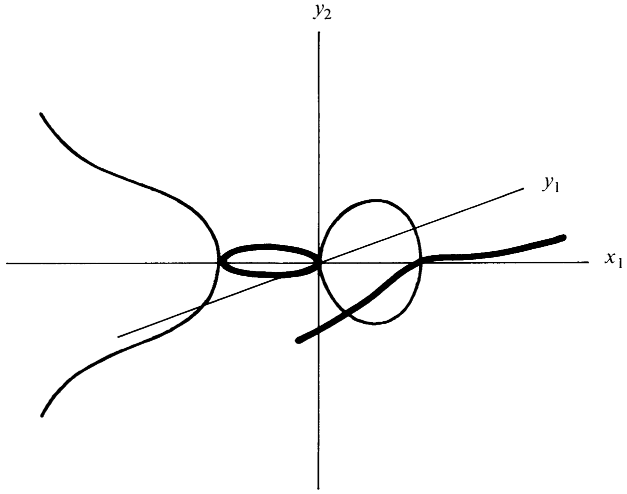


FIG. 2

The two branches on the right and left meet in the point at infinity P_∞ which was added to $\mathbb{C}_y$ (since $|y|/|x| \to \infty$ as $|x| \to \infty$). We now have three topological loops, each touching the other two. These can be bent to lie on a torus (Fig. 3); it turns out that the other curves corresponding to $x_2 \neq 0$ fill in the rest of the torus. Thus, of the whole torus, all we see in the usual real part are the two dark loops in Fig. 3, without P_∞.

One can similarly show, for instance, that

$$y^2 - x(x^2 - 1^2)(x^2 - 2^2) \cdots (x^2 - g^2)$$

defines in $\mathbb{P}^2(\mathbb{C})$ a closed orientable surface of genus g.

Through these examples, we see that doing analytic geometry in $\mathbb{P}^2(\mathbb{C})$ on the one hand reveals a fuller picture of curves defined by polynomials, and on the other hand leads to an appreciation that orientable topological surfaces—donuts, coffee cups, pretzels and all—aren't just "made up," but arise in a very natural way. For further details and pictures on "algebraic curve sketching" over $\mathbb{C}$, see [7].

6. A Pearl from Algebraic Geometry. We promised earlier that our new viewpoint would lead us to a much more general form of the Fundamental Theorem of Algebra. To see how, let's briefly consider once again the first example of the parabola $y = x^2$, looked at in the complex setting. The Fundamental Theorem of Algebra guarantees that the complex line $\mathbb{C}_x$ intersects the parabola in two points, counted with multiplicity. "Multiplicity two" means, intuitively, that if we translate $\mathbb{C}_x$ a little into a new line $\mathbb{C} \neq \mathbb{C}_x$, then the point of intersection separates into two

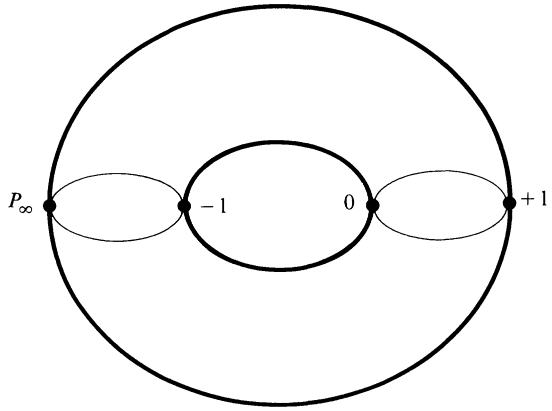


FIG. 3

distinct points, which then coalesce as $\mathbb{C}$ returns to its original position $\mathbb{C}_x$. (This idea can be generalized almost beyond recognition.) More generally, one can show, using a linear change of coordinates, that any complex line of the form $y = ax + b$ $(a, b \in \mathbb{C})$ intersects the parabola in two points. *Question*: Does *every* complex line in $\mathbb{P}^2(\mathbb{C})$ intersect it in two points? *Answer*: Yes. For the branches of the parabola approach a point at infinity, P_∞—the one added to $\mathbb{C}_y$; it can then be shown that any line in $\mathbb{P}^2(\mathbb{C})$ defined by $x = a$ intersects the completed parabola in two points (at (a, a^2) in $\mathbb{C}_{xy}$, and at P_∞). The only line left to consider is the one at infinity. Looking at it as the limit of, say, $y = c$ as $|c| \to \infty$, one can see that the line at infinity and the parabola intersect at P_∞ with multiplicity two. (Relative to the topology we put on $\mathbb{P}^2(\mathbb{C})$, the two points of intersection of $y = c$ with the parabola coalesce to P_∞ as $|c| \to \infty$.)

Note that the degree of a parabola is two, the degree of a line is one, and that $2 \cdot 1 =$ the number of points of intersection. The general form of the Fundamental Theorem of Algebra we promised is "Bézout's theorem." Before stating it we need only remark that the idea of determining the multiplicity of intersection by translating one of the curves a bit, generalizes to arbitrary curves C_1, C_2 in $\mathbb{P}^2(\mathbb{C})$ defined by nonconstant polynomials $q_1(x, y)$ and $q_2(x, y)$. This may be done as follows. Let an open neighborhood U in some $\mathbb{C}^r$ parametrize the set of all "small" changes in the (nonzero) coefficients of q_1. Then for any isolated point P in $C_1 \cap C_2$, there is a nowhere dense subset of U off which this holds: The corresponding curve C_1' intersects C_2, near P, in exactly m points. We then say *C_1 and C_2 intersect at P with multiplicity m.*

BÉZOUT'S THEOREM. *Let q_1, q_2 be distinct, irreducible polynomials in $\mathbb{C}[x, y] \setminus \mathbb{C}$; let them define curves C_1, C_2 in $\mathbb{P}^2(\mathbb{C})$. Then C_1 and C_2 intersect in* $(\deg q_1) \cdot (\deg q_2)$ *points, counted with multiplicity.*

REMARKS:

(1) In the Fundamental Theorem of Algebra, q_1 and q_2 are very special: q_1 is of the form $y - p(x)$—i.e., it defines the graph of a function $\mathbb{C}_x \to \mathbb{C}_y$; and q_2 is y. There is no need for such restrictions in Bézout's Theorem. Both curves are put on an equal and more general footing.

(2) One can easily draw in $\mathbb{R}^2$ two ellipses intersecting in $2 \cdot 2 = 4$ points. As one ellipse is translated with respect to the other, we begin to "lose" points of intersection in $\mathbb{R}^2$; but not in $\mathbb{P}^2(\mathbb{C})$—the "lost" points simply have become complex. Thus two complex ellipses, disjoint in $\mathbb{R}_{xy}$, extend to curves in $\mathbb{P}^2(\mathbb{C})$ having four complex points of intersection.

(3) Bézout's Theorem can also be stated for polynomials which are not necessarily irreducible: if neither q_1 nor q_2 have repeated factors, and if q_1, q_2 share no nonconstant factors, then C_1 and C_2 intersect in $(\deg q_1) \cdot (\deg q_2)$ points.

Bézout's Theorem is named after the French mathematician Etienne Bézout (1730–1783). Although his main occupation was mathematics examiner in the French Naval Academy (and later in the Royal Artillery), he wrote two textbooks which became so popular they were translated into several languages and ran many editions. Bézout, who worked on "elimination theory" using linear equations, stated his theorem in terms of a determinant derived from the coefficients of q_1 and q_2, called the "resultant" of q_1 and q_2.

7. Some Additional Comments on Intersection Multiplicity. Bézout's Theorem hinges on the concept of intersection multiplicity. This concept turns out to be an excellent example of the "three blind men touching an elephant" phenomenon in mathematics—one big idea (the elephant) may appear to be very different to different people (and different generations). For instance, our definition of intersection multiplicity above looks almost totally geometric. In this section we give for curves a second and a third definition; one of them appears very algebraic, the other, purely topological. Yet over $\mathbb{C}$, all definitions are equivalent and each one provides a basis for generalizing in a different direction.

For the algebraic approach we start with the example of the complex parabola $y - x^2 = 0$ intersecting $\mathbb{C}_x$. Now $\mathbb{C}_x$ may be parametrized about $(0,0)$ by $x = t$, $y = 0$. Substituting into $y - x^2 = 0$ gives t^2. The order of the polynomial t^2 is two, which is just the intersection multiplicity of $y - x^2$ with $\mathbb{C}_x$! Any other complex line L through $(0,0)$ intersects the parabola there in multiplicity one. In agreement with this, L may be parametrized by $x = at$, $y = bt$ $(b \neq 0)$; substituting this into $y - x^2$ gives $bt - a^2t^2$ which has order one at 0.

One could just as well substitute a parametrization of the parabola into the line. Parametrize $y = x^2$ by $x = t$, $y = t^2$, and plug into $Ax + By$. We get $At + Bt^2$ which has order one except when $A = 0$ (that is, except when L is $\mathbb{C}_x$); when $A = 0$ the order is two. As another example, consider two parabolas tangent at $(0,0)$—say, $y - x^2 = 0$ and $y - 4x^2 = 0$. Substituting $x = t$, $y = t^2$ into $y - 4x^2$ gives $-3t^2$, so the parabolas intersect in a double point at $(0,0)$.

The general idea here is this: to find the intersection multiplicity m_P of two curves C_1 and C_2 at P, parametrize one curve about P, and substitute these parametric equations into the equation of the other curve. The resulting order at zero is m_P. *A caveat*: use only "reduced" parametrizations. Thus $\mathbb{C}_x$ *could* be parametrized about $P = (0,0)$ by $x = t^2$, $y = 0$ (or the cusp curve $y^2 = x^3$ by $x = t^4$, $y = t^6$). But for correct m_P, first reduce the parametrizations to lowest terms, getting $x = t$, $y = 0$ (or $x = t^2$, $y = t^3$).

It turns out that parametrizations can be found for any algebraic curve about any point, though at a "cross point" or "singularity," one may need more than one parametrization there. For example, the curve $y^2 = x^2(x + 1)$ crosses itself at $(0,0)$ (looking like an "alpha" in $\mathbb{R}_{x_1y_2}$). Each part, or "branch" has its own parametrization. (Such parametrizations at singular points may be power series, but that doesn't matter—the concept of order is still the same.) One then just adds up the multiplicities due to the different branches through the point. An elementary proof of the existence of such parametrizations, which includes an explicit method for finding them, is given in [**9**, Chap. IV].

Now let's turn to the topological viewpoint. Let S_ε^3 denote the three-sphere $x_1^2 + x_2^2 + y_1^2 + y_2^2 = \varepsilon^2$ in $\mathbb{C}_{xy}$ $(= \mathbb{R}_{x_1x_2y_1y_2})$; $\varepsilon > 0$ is assumed small. S_ε^3 intersects $\mathbb{C}_x$ in a real circle C, and S_ε^3 intersects $y = x^2$ in a topological circle which winds around C—twice! The intersections of S_ε^3 with $\mathbb{C}_x$ and $\mathbb{C}_y$ have mutual linking number *one*; and the intersections of S_ε^3 with $y = x^2$ and $y = 4x^2$ have linking number two. For a curve like $y^4 = x^9$, for all sufficiently small ε, $S_\varepsilon^3 \cap (y^4 = x^9)$ winds around $S_\varepsilon^3 \cap C_x$ nine times. But if L is any other complex line through $(0,0)$, then for all sufficiently small ε, $S_\varepsilon^3 \cap (y^4 = x^9)$ winds around $S_\varepsilon^3 \cap L$ exactly four times. The reader can easily check that these linking numbers agree with the multiplicities obtained using the substitution method above.

8. The World's Most Elusive Curve-Sketching Problem. We have seen how, if we replace $\mathbb{R}$ by $\mathbb{C}$, analytic geometry leads us in a natural way to topology. But $\mathbb{R}$ may be replaced by other fields, too. For instance, doing analytic geometry over $\mathbb{Q}$ instead of over $\mathbb{R}$ leads directly to number

theory. To see this we will try sketching, in $\mathbb{Q}^2$, some curves defined by polynomials with coefficients in $\mathbb{Q}$.

As a first example, consider the line L in $\mathbb{Q}^2$ defined by $ax + by + c = 0$ $(a, b, c \in \mathbb{Q})$. It looks like the corresponding line L' in $\mathbb{R}^2$, in the sense that L is everywhere dense in L'. It is also easily checked that the same thing is true of the "parabola" in $\mathbb{Q}^2$ defined by $y = x^2$—it is dense in the graph of $y = x^2$ in $\mathbb{R}^2$.

How about the "rational circle" $x^2 + y^2 = 1$ in $\mathbb{Q}^2$? Does it look like the corresponding real circle? This time the answer isn't so obvious, and we are already led to an interesting question in number theory. What we are asking is this: for any point P on the real circle, is there a point $P' \in \mathbb{Q}^2$ arbitrarily near P and on that circle? We begin by observing that if (a, b) is any point of the rational circle ($(3/5, 4/5)$ is an example), and if c is a common denominator of a and b, then ac, bc and c are integers satisfying the equation $(ac)^2 + (bc)^2 = c^2$. Put another way, (ac, bc) is a "Pythagorean point" in $\mathbb{R}^2$—a point (p, q) with integer co-ordinates so that $(0,0)$, $(p,0)$, (p, q) are the vertices of a triangle having sides of integral length. (Thus $(3,4)$ is a Pythagorean point in $\mathbb{R}^2$.) We may now rephrase the question this way: *Are the slopes m of those 1-subspaces of $\mathbb{R}^2$ passing through Pythagorean points dense in $\mathbb{R}$?* As it happens, all Pythagorean points are known (this is the number theory part): any point $(p^2 - q^2, 2pq)$ (p, q distinct, nonzero integers) is Pythagorean, and any Pythagorean point is either of this form, or an integral scalar multiple of it. One can now easily verify that for any $m \in \mathbb{R}$ and $\varepsilon > 0$, there is some $2pq/(p^2 - q^2)$ within ε of m. So the rational circle $x^2 + y^2 = 1$ is indeed dense in the real one.

What about other curves—for example $x^n + y^n = 1$? In $\mathbb{R}^2$, for $n > 2$ and even, it looks like something like the perimeter of a TV screen, becoming more and more "square" as n gets larger. For $n > 2$ and odd, it has two branches asymptotic to $y = -x$. Does an appropriate generalization of the above show that the corresponding rational curve is dense in the real one? No, it doesn't. In fact it has been conjectured, though never proved, that except for "trivial points" (in which one co-ordinate is zero), in $\mathbb{Q}^2$ the curve is actually empty! Does this sound familiar? It is a geometric way of looking at "Fermat's Last Theorem": for $n > 2$, $a^n + b^n = c^n$ has no nontrivial integer solutions. For, dividing through by c^n, we can restate it as: *if $n > 2$, then $r^n + s^n = 1$ has no nontrivial rational solutions.* So Fermat's Last Theorem is just a (very elusive) curve-sketching problem in $\mathbb{Q}^2$.

In $\mathbb{P}^2(\mathbb{C})$, we found that irreducible conics are topologically spheres. We can similarly try to sketch conics in $\mathbb{Q}^2$ defined by $ax^2 + bxy + cy^2 + dx + ey + f = 0$ $(a, \ldots, f \in \mathbb{Q})$. We've already found that the rational unit circle looks like the corresponding real circle. *Question*: Is the analogous thing true for the above "rational conics"? *Answer*: Not at all; for some choices of coefficients the rational locus is empty, for others there are finitely many solutions, and for still others it is dense in the real sketch. A full answer to this question is highly nontrivial. In fact, a large part of Gauss' famous *Disquisitiones arithmeticae* is devoted to answering this and other questions closely related to it. It should be mentioned that one often extends $\mathbb{Q}$ to a finite algebraic extension of it, or replaces $\mathbb{Q}$ by a finite field and asks for the set of solutions to such Diophantine problems over the new field.

9. Algebra: Old versus New. So far we've considered mostly what can be described as "curves in 2-spaces." Each such curve can be defined using a single polynomial. However, a single polynomial will in general define different sets in different spaces. Thus x defines a point in $\mathbb{C}_x$, a line in $\mathbb{C}_{xy}$, a plane in $\mathbb{C}_{xyx}$ and a hyperplane in $\mathbb{C}^n$. In each case the dimension is $n - 1$. It turns out that more generally, any nonconstant polynomial $p \in \mathbb{C}[x_1, \ldots, x_n]$ defines a "hypersurface" in $\mathbb{C}_{x_1, \ldots, x_n}$ (that is, a surface of complex dimension $n - 1$). This is a basic extension of the case $n = 2$, where we saw that even polynomials like $x^2 + y^2 - 1$ (whose set of zeros in $\mathbb{R}^2$ is empty) continues to define a full-fledged complex curve (have two real dimensions) in $\mathbb{C}^2$.

But this fact can be looked at from the other end—namely, nonconstant single polynomials can yield *only* hypersurfaces in $\mathbb{C}^n$, nothing any different. It means that restricting oneself to single polynomials has serious limitations. We can conveniently see the situation in $\mathbb{R}^3$. Suppose,

for example, we try to define a real circle in $\mathbb{R}^3$. We soon find out that *two* polynomials are needed; roughly, each polynomial "cuts down the dimension by one." Thus we can look at the circle as the intersection of: a plane and a cylinder; or a plane and a sphere; or a paraboloid and a cone, and so on. There is no single "canonical" pair defining the circle. The most natural thing to do is to consider the set of *all* polynomials vanishing on the circle. This forms an *ideal* in the ring of all polynomials in three variables. This idea holds equally well over other fields—in particular, over $\mathbb{C}$ or $\mathbb{Q}$. We call the set of zeros common to any collection of polynomials an *algebraic variety*, or, for short, a *variety*. Not only does a variety define an ideal, but any ideal of course defines a variety; we will find that varieties (geometry) and ideals (algebra) naturally link up, forming a kind of "dictionary." Thus in extending analytic geometry from plane curves to varieties in higher dimensions, we are led from "old" (or high school) algebra to a part of "modern" algebra, namely commutative ring theory. And just as we experienced some enlightenment in sketching curves in $\mathbb{P}^2(\mathbb{C})$ for instance, one finds that old facts from ring theory take on new geometric meaning; in fact, one can often predict, on the basis of geometric intuition, which statements in commutative ring theory are likely to be true and which are not. We indicate only a few highlights here, but they will supply a bit of the favor.

For concreteness, we will consider varieties in $\mathbb{C}^n$ defined by collections of polynomials chosen from the ring $R = \mathbb{C}[x_1,\dots, x_n]$. If $\mathfrak{a}$ is an ideal of R, we will let $V(\mathfrak{a})$ denote the common set of zeros of all the polynomials in $\mathfrak{a}$; this is called the *variety defined by* $\mathfrak{a}$.

Recall that in $\mathbb{C}_{xy}$, $x^2 + y^2$ defines the union of two complex lines, since a point is a zero of $x^2 + y^2$ $(= (x + iy)(x - iy))$ iff it makes at least one of $x + iy$, $x - iy$ zero. More generally, for polynomials p and q the set of zeros of the product pq is the union of the zero sets of p and of q. This generalizes to the ideal-theoretic level, as follows. First, if $\mathfrak{a}$ and $\mathfrak{b}$ are ideals of R, then the *product* $\mathfrak{a}\mathfrak{b}$ is the set of all sums of products ab ($a \in \mathfrak{a}$, $b \in \mathfrak{b}$); $\mathfrak{a}\mathfrak{b}$ is an ideal of R.

Fact: $V(\mathfrak{a}\mathfrak{b}) = V(\mathfrak{a}) \cup V(\mathfrak{b})$.

With ideals, we get even more than with polynomials: If we define $\mathfrak{a} + \mathfrak{b}$ to be the set of sums $a + b$ ($a \in \mathfrak{a}$, $b \in \mathfrak{b}$), then $\mathfrak{a} + \mathfrak{b}$ is an ideal and we have this

Fact: $V(\mathfrak{a} + \mathfrak{b}) = V(\mathfrak{a}) \cap V(\mathfrak{b})$.

It can be shown, using the above two facts, that the union and intersection of any two varieties is again a variety. The operation "V" sets up a basic link between algebraic geometry and commutative ring theory.

If ideals in fact generalize polynomials, what ideal-theoretic concept generalizes "irreducible polynomial"? The answer is *prime ideal.* (Recall that an ideal $\mathfrak{p}$ is prime if $p = qr \in \mathfrak{p}$ implies that $q \in \mathfrak{p}$ or $r \in \mathfrak{p}$.) It is easy to check that if p is irreducible, than the ideal (p) is prime. At the geometric level, we say that an algebraic variety is *irreducible* if it isn't the union of two property smaller algebraic varieties. The following hooks up these two ideas:

Fact: If $\mathfrak{p}$ is prime, then $V(\mathfrak{p})$ is irreducible.

Often an ideal can be written as a product of prime ideals, and then one has:

Fact: A splitting of $\mathfrak{p}$ into prime ideals corresponds to a decomposition of $V(\mathfrak{p})$ into irreducibles; that is, if $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$, then $V(\mathfrak{a}) = V(\mathfrak{p}_1) \cup \cdots \cup V(\mathfrak{p}_r)$.

It is interesting to note that although $\mathfrak{p}$ prime implies $V(\mathfrak{p})$ irreducible, the converse is not in general true— $V(\mathfrak{a})$ can be irreducible without being prime.

EXAMPLE: $\mathfrak{a} = (y, y - x^2) \subseteq \mathbb{C}[x, y]$ is not prime since, e.g., $y - (y - x^2) = x^2 = x \cdot x$ is in $\mathfrak{a}$, but x is not. But $V(\mathfrak{a}) = (0,0) \in \mathbb{C}_{xy}$ since only $(0,0)$ is common to $y = 0$ $(= \mathbb{C}_x)$ and $y - x^2 = 0$ (a parabola); and $(0,0)$ is surely irreducible.

Fact: Many ideals can define the same algebraic variety.

Now if there really is a good "dictionary," then aren't the different ideals defining the same variety trying to tell us something more? From the definition of algebraic variety, it is easy to see that if $\mathfrak{a} \subseteq \mathfrak{b}$, then $V(\mathfrak{a}) \supseteq V(\mathfrak{b})$; that is, V is a partial-order reversing map from ideals to varieties. Now if $V(\mathfrak{a}) = V(\mathfrak{b})$, with $\mathfrak{a} \subsetneqq \mathfrak{b}$, then isn't $\mathfrak{a}$ somehow defining something "larger" than $\mathfrak{b}$ is?

In the example of $\mathfrak{a} = (y, y - x^2)$ above, $V(\mathfrak{a}) = (0,0)$. But if $\mathfrak{b} = (x, y)$, then $V(\mathfrak{b}) = (0,0)$ too, and it is not hard to check that $\mathfrak{a} \subsetneqq \mathfrak{b}$. Geometrically, $V(\mathfrak{b})$ may be looked at as the intersection of $x = 0$ (i.e., $\mathbb{C}_y$) and $y = 0$ (i.e., $\mathbb{C}_x$). But $V(\mathfrak{a})$ is obtained by intersecting $y = 0$. Also, $y = x^2$ (a parabola), and the parabola intersects $\mathbb{C}_x$ with multiplicity *two*! Since $\mathbb{C}_x$ and $\mathbb{C}_y$ intersect in only multiplicity *one*, $\mathfrak{a}$ in a sense defines an object (double point) larger than $\mathfrak{b}$ does (single point).

Fact: Ideals not only define varieties; they also supply information about multiplicities.

A generation ago, men such as Chevalley, Weil and Zariski did much to make such ideas precise, thus cementing the relationship between algebraic geometry and commutative ring theory. This has been continued in much expanded contexts by Grothendieck and others.

10. The Fundamental Theorem of Algebra Revisited. We present here two famous generalizations of the Fundamental Theorem of Algebra to higher dimensions; the statement of one is more algebraic in flavor and the other, more geometric.

We consider the algebraic one first. Let us begin by noting that since $\mathbb{C}[x]$ is a principle ideal domain there is, up to units, a natural one-to-one correspondence between the elements and the ideals of $\mathbb{C}[x]$: every element p in $\mathbb{C}[x]$ generates the principal ideal (p); and conversely, every ideal of $\mathbb{C}[x]$ is of the form (p), where p is uniquely determined up to a nonzero constant of $\mathbb{C}[x]$. Note that any constant of $\mathbb{C}[x]$ generates the ideal (0) or $\mathbb{C}[x]$; any other element of $\mathbb{C}[x]$ generates a proper ideal of $\mathbb{C}[x]$.

Next, in analogy with the notion of "zero of a polynomial," we may, for any ideal $\mathfrak{a} \subseteq \mathbb{C}[x_1,\dots, x_n]$, define a *zero of* $\mathfrak{a}$ to be any point in $\mathbb{C}^n$ at which every polynomial of $\mathfrak{a}$ vanishes. (Thus $V(\mathfrak{a})$ is simply the set of all zeros of $\mathfrak{a}$.)

Now the Fundamental Theorem of Algebra—"every nonconstant polynomial in $\mathbb{C}[x]$ has a zero in $\mathbb{C}$"—can be translated to the language of ideals: *Every proper ideal of* $\mathbb{C}[x]$ *has a zero.* This italicized statement generalizes directly to ideals in $\mathbb{C}[x_1,\dots, x_n]$; it is the celebrated

Hilbert Nullstellensatz: Every proper ideal of $\mathbb{C}[x_1,\dots, x_n]$ has a zero.

It tells us that for any ideal $\mathfrak{a} \subsetneqq \mathbb{C}[x_1,\dots, x_n]$, $V(\mathfrak{a})$ is nonempty. There are various generalizations and analogues of this theorem. For instance, given any proper *principal* ideal $\mathfrak{a} = (p) \subsetneqq \mathbb{C}[x_1,\dots, x_n]$, it happens that $V(\mathfrak{a})$ is not only nonempty, but that its complex dimension in $\mathbb{C}^n$ is $n - 1$ (that is, real dimension $2n - 2$). This extends what we found out earlier about curves defined by nonconstant polynomials in $\mathbb{C}[x, y]$: the full picture there has complex dimension $2 - 1 = 1$, or real dimension 2. Very roughly, the further away a proper ideal $\mathfrak{a}$ is from being principal, the smaller the dimension of $V(\mathfrak{a})$.

Hilbert's Nullstellensatz also generalizes from $\mathbb{C}$ to any algebraically closed field $\mathbb{K}$; the varieties $V(\mathfrak{a})$ are then taken to lie in $\mathbb{K}^n$.

We now consider our second generalization of the Fundamental Theorem of Algebra. First, generalizing the transition from $\mathbb{C}^2$ to $\mathbb{P}^2(\mathbb{C})$, one can add a point at infinity to the set of complex lines parallel to each 1-subspace of $\mathbb{C}^n$, obtaining $\mathbb{P}^n(\mathbb{C})$; $\mathbb{P}^n(\mathbb{C})$ may then be supplied with a topology induced from the 1-subspaces of $\mathbb{C}^{n+1}$ passing through $\mathbb{C}^{n+1}$-open subsets, as we analogously did with $\mathbb{P}^2(\mathbb{C})$. Any variety in $\mathbb{C}^n$ may then be completed in $\mathbb{P}^n(\mathbb{C})$ by taking its topological closure.

Now Bézout's Theorem, already an important generalization of the Fundamental Theorem of Algebra, in fact extends to higher dimensions, and this is our second generalization. To explain, for any variety V in $\mathbb{P}^n(\mathbb{C})$ it turns out that there is some $m \leqslant n$ such that "almost all" subspaces

$\mathbb{P}^m(\mathbb{C})$ of $\mathbb{P}^n(\mathbb{C})$ intersect V in a fixed number $d > 0$ distinct points. (All subspaces $\mathbb{P}^m(\mathbb{C})$ of $\mathbb{P}^n(\mathbb{C})$ may be parametrized in a natural way by an algebraic variety G (a "Grassmann manifold"); "almost all" means all subspaces corresponding to points off a proper algebraic subvariety of G.) V is then said to have *dimension* $n - m$ (or *codimension* m), and *degree* d. As with curves, a variety is said to be irreducible if it isn't the union of any two strictly smaller varieties.

GENERALIZED BÉZOUT THEOREM: *Let V, W be two distinct irreducible algebraic varieties in $\mathbb{P}^n(\mathbb{C})$ of codimensions r, s and degrees d, e. If $r + s \leqslant n$, then $V \cdot W$ has codimension $r + s$ and degree de.*

REMARKS:

(1) We've written $V \cdot W$ instead of $V \cap W$ to indicate that the intersection is "counted with multiplicity." A geometric definition of multiplicity may be given in the same spirit as for curves. The following example (which may be visualized in $\mathbb{R}_{xyz}$ although it actually takes place in $\mathbb{C}_{xyz}$) gives the idea. The cylinder $x^2 + y^2 = 1$ (codimension 1, degree 2) tangentially intersects the sphere $x^2 + y^2 + z^2 = 1$ in a *double* circle—that is, a circle counted with multiplicity two. (The double circle has codimension $1 + 1 = 2$ and degree $2 \cdot 2 = 4$.) Almost all small changes in the coefficients of our two polynomials will separate the points of the double circle. For instance, shrinking the cylinder a little bit will split up the double circle into two separate circles. The cylinder and sphere now intersect "transversally," and one can easily check that the union of the two circles has degree four; this union has multiplicity one, in that additional small "algebraic perturbations" of the sphere or cylinder won't further separate points.

As with curves, intersection multiplicity may also be looked at in a more algebraic way, as well as in a more topological fashion. One of the important milestones in algebraic geometry is Weil's book [**10**], in which he works out the theory of intersection multiplicity in a purely algebraic way for algebraic varieties of arbitrary dimension over arbitrary groundfield.

(2) The theorem can be stated to accommodate general, not-necessarily-irreducible algebraic varieties.

(3) The theorem can be generalized to any algebraically closed field $\mathbb{K}$, the varieties lying in $\mathbb{P}^n(\mathbb{K})$.

Before we leave the world of higher dimensions, we make a final observation. Just as one can do curve sketching over "arithmetic" fields $\mathbb{K}$ such as $\mathbb{Q}$, algebraic extensions of $\mathbb{Q}$ and finite fields, one can ask for the appearance in, say, $\mathbb{K}^n$ of varieties defined by polynomials with coefficients in $\mathbb{K}$. We are lead to even more far reaching results (as well as interesting new questions) in number theory. For example, a study of the solution in $\mathbb{K}^n$ of a quadratic form in n variables brings us all the way from the "$n = 2$" of Gauss' *Disquisitiones arithmeticae* (1801) through deep studies by Hilbert, Minkowsky, Hasse, C. L. Siegel and others. Much work on these number-theoretic questions continues to this day.

11. What About Calculus? A major sales pitch of this article has been that it is enlightening to study analytic geometry over fields other than $\mathbb{R}$. One might wonder if the same is true for calculus. The answer is yes. Algebraic geometry enters the picture in a natural way; we briefly indicate how, in this section.

We replace $\mathbb{R}$ by $\mathbb{C}$. The reader is probably familiar with integration over $\mathbb{C}$, in which "line integrals" appear. In studying this, one soon meets a problem such as integrating $\int\sqrt{x}\,dx$ around $0 \in \mathbb{C}_x$. There we encounter the multiple-valuedness of $\sqrt{x}$, and we find that we must go around the origin twice to arrive back at the original value of $\sqrt{x}$. In this sense, then, it is somewhat artificial to insist that we're integrating on $\mathbb{C}$; the integration is more naturally taking place on a Riemann surface which, in our context, amounts to an algebraic curve which is "nonsingular." (Nonsingular means that for each point P of the curve, in some neighborhood of P the curve is the graph of an invertible complex-analytic function. Thus the "circle" $x^2 + y^2 = 1$ in $\mathbb{C}_{xy}$ is

nonsingular; but $\mathbb{C}_x \cup \mathbb{C}_y$ is not, since the part around $(0,0)$ isn't the graph of a function.) Let us explain. In $\mathbb{C}_{xy}$ the curve C given by $y^2 = x$ has a function on it, namely the restriction to C of the coordinate function y on $\mathbb{C}_{xy}$, and we are integrating around a closed path on C. Of course one can integrate other functions on C, too—for instance (the restriction to C of) rational functions of x and y.

Note: Instead of looking at C as the union of the slices $x_2 =$ constant used earlier, we could just as well let our slices be cylinders (real circle about $(0,0)$ in $\mathbb{C}_x) \times \mathbb{C}_y$. Now the intersection of each such cylinder with C is a simple, closed curve lying over the real circle. As a point goes around the closed curve once, its projection on the circle goes around twice; the whole complex parabola is the union of $(0,0)$ and all these "two loop" curves, one such curve for each of the concentric cylinders. We can now think of integrating around $(0,0)$ as actually integrating along one of these curves.

Now consider the line integral $\int\sqrt{x(x-1)(x+1)}\,dx$. We once again encounter multiple-valuedness when integrating around a simple closed curve in $\mathbb{C}_x$ containing an odd number of the points $0, 1, -1$. Again it is natural to look at this as integration of y on the curve $y^2 = x(x-1)(x+1)$. In $\mathbb{P}^2(\mathbb{C})$, this defines the torus of §5; now it serves as a "Riemann surface of genus one." More generally, the integral $\int\sqrt{x(x^2-1^2)\cdots(x^2-g^2)}\,dx$ leads to a Riemann surface of genus g.

What can be said about the values obtained by integrating on such curves? First, recall that in $\mathbb{C}$, if a function is analytic in a simply-connected region R, then for fixed points Q_0, Q_1, the value of any line integral $\int_{Q_0}^{Q_1} f(x)\,dx$ is independent of the path in R connecting Q_0 and Q_1. But if f has, say, one pole P in R, then the value depends on the number of circuits the path makes around P. One can regard P as a "hole," in the sense that f is analytic in $R \setminus P$ but not in R. This time any two values of $\int_{Q_0}^{Q_1} f(x)\,dx$ differ by an integral multiple of a period (which is the value of $\int_\gamma f(x)\,dx$, where γ is closed and goes around P once).

What happens when we integrate on, say, a nonsingular curve of genus *one*? The situation is analogous to the above, only now there are *two* "holes" since a torus is topologically the product circle $\times$ circle, and each of the two circles acts like a hole. Thus, the values of $\int_{Q_0}^{Q_1} f(x)\,dx$ (Q_0, Q_1 fixed) form a point lattice in $\mathbb{C}$. If one holds Q_0 fixed but lets Q_1 vary, the values that $\int_{Q_0}^{Q_1} f(x)\,dx$ attains fill out all of $\mathbb{C}$. If the path of integration never topologically encircles a hole of the torus, the values fill out only a part of $\mathbb{C}$—namely, a "fundamental region," a parallelogram topologically equal to $[0,1) \times [0,1)$, containing the point $0 \in \mathbb{C}_x$. Note that the set of values modulo the periods, i.e., $\mathbb{C}/(\mathbb{Z} \times \mathbb{Z})$, is a torus. One can actually give this the structure not only of a commutative group, but also of an algebraic variety in $\mathbb{P}^2(\mathbb{C})$, obtaining an "abelian variety."

What happens when we integrate on a nonsingular curve of genus g? Now there are $2g$ holes, the values of $\int_{Q_0}^{Q_1} f(x)\,dx$ (Q_0 fixed, Q_1 varying) fill out $\mathbb{C}^g$, and the values of $\int_{Q_0}^{Q_1} f(x)\,dx$ (Q_0, Q_1 both fixed) fill out a lattice in $\mathbb{C}^g$ looking like $\mathbb{Z}^{2g}$. If the path γ connecting Q_0 and Q_1 never encircles a hole, the values form a fundamental region looking like $[0,1)^{2g}$; and $\mathbb{C}^g/\mathbb{Z}^{2g}$ is a product of $2g$ real circles, which is an abelian variety called the "Jacobian variety" of the curve.

It should be mentioned that the restriction of nonsingularity is not actually necessary. Any curve in $\mathbb{P}^2(\mathbb{C})$ can be "desingularized;" for us, this amounts to replacing it by a nonsingular curve having all the function-theoretic information of the original curve.

These ideas have been extended to varieties of higher dimension, and much effort has gone into trying to extend as much of all this as possible to the "arithmetic" fields mentioned earlier.

12. Newer Methods in Algebraic Geometry. The "evolution of science" is often more like a series of revolutions. In mathematics the introduction of imaginary quantities, of differentials, of transfinite numbers, of nonconstructive existence proofs—the list could go on and on—all represented revolutions in their day. Riemann began a revolution with his introduction of Riemann surfaces, and it profoundly affected algebraic geometry. Hilbert's nonconstructive proof in 1888 of the Nullstellensatz (§10) was wildly new then, prompting P. Gordan, a leading algebraic

geometer of the day, to his famous protest "Das ist nicht mathematische—das ist theologie!" More recently the school including Weil, Chevalley and Zariski revolutionized algebraic geometry by breaking away from the constraints of working over $\mathbb{C}$, replacing $\mathbb{C}$ by an arbitrary groundfield; this "arithmetization of algebraic geometry" helped bring large parts of number theory into the fold of algebraic geometry.

In each case, it was not a matter of a change in fashion for fashion's sake; the power of any cluster of methods to solve new problems gradually gets exhausted, meeting the point of diminishing returns—and new approaches and concepts are needed.

The 1950's found a new storm brewing, one that ended up absorbing into algebraic geometry nearly all of commutative algebra, and topological notions such as fiber bundles, sheaves and various cohomological theories. In this final section we briefly indicate some of these newer ideas as they touch on algebraic geometry. An excellent introductory treatment can be found in [**4**].

By way of motivation, let us begin with an irreducible variety $V \in \mathbb{C}^n$ and let $\bar{p}$ denote the restriction to V of the polynomial $p \in \mathbb{C}[x_1, \ldots, x_n]$. It turns out that all such $\bar{p}$ form an integral domain D (no zero divisors); let K be its quotient field. We can regard elements of K as rational functions on the variety. (K is called the function field of V.) We can call the elements of the integral domain D "integral functions" on V. An element algebraic over K is an algebraic function on V. The similarity in nomenclature with numbers is striking—one has integral, rational and algebraic numbers. Nonalgebraic numbers are called transcendental, which are analogous to "transcendental functions." Is this similarity accidental, or is something fundamental going on? We can ask a leading question right off: can we actually look at the ordinary integers $\mathbb{Z}$ and their quotients $\mathbb{Q}$ as functions on some "variety"? The answer is yes. We briefly explain. First, all functions of D vanishing at a point $P \in V$ form a maximal ideal $\mathfrak{m}_P$ in D. (A point is a smallest nonempty object in V; a maximal ideal is a largest $\neq D$ ideal in D.) It is not hard to see that for any function in D, its value at P is its ring-homomorphism image in $D/\mathfrak{m}_P = \mathbb{C}$. It turns out that every maximal ideal of D is some $\mathfrak{m}_P$ and that P and $\mathfrak{m}_P$ are in $1:1$ correspondence. We may thus identify the set of points in V with the set of maximal ideals of D.

What variety has $\mathbb{Z}$ as "D"? In $\mathbb{Z}$ the maximal ideals are just the nonzero prime ideals. Thus the "variety" having $\mathbb{Z}$ as "ring of integral functions" has for its points the set of prime ideals (2), (3), (5),... ! The elements of $\mathbb{Z}$ may now be looked at as functions on our "variety." For example 25 at the point (p) is 25 mod(p); therefore it is 1 at (2), 1 at (3), 0 at (5), 4 at (7), and so on. The elements of Q are rational functions; $3/49$, for example, is "regular" except for a pole of order 2 at the point (7).

It soon became clear to algebraic geometers that one could profitably be very general and consider *any* commutative ring R with unit in place of D. One could furthermore consider not only maximal ideals as points, but prime ideals (which generalize maximal ideals) as more general points. These more general points are like irreducible varieties of V which in many respects behave like ordinary points. The set of all prime ideals of R is denoted "Spec R", and is suitably topologized. One can then build on Spec R to get an "affine scheme": roughly speaking, it is Spec R together with, at each $P \in \operatorname{Spec}(R)$, the functions (quotients of R) regular in some neighborhood P. One can glue these affine schemes together to get more general schemes, much as one glues together topological, differentiable or analytic neighborhoods to construct more general topological, differentiable or analytic manifolds.

But another current was to make its way into algebraic geometry. Workers in manifold theory had been experimenting with fiber bundles, sheaves, various cohomological theories and the like. These powerful tools were yielding subtle invariants of the manifolds and could be used, for instance, to help classify those manifolds. It was realized that many of these ideas could be extended to schemes, and could provide to algebraic geometry new, useful invariants. However, all this ended up (1) being pretty abstract, and (2) introducing a lot of new, heavy baggage, baggage to be added to a subject already heavy with structure. The big question: is it worth it? Now, after it has had time to prove itself, we can say definitely yes. It has been all-important in solving many

"big" problems in algebraic geometry, some of long standing. Occasionally, after a modern proof has been found, a more elementary one follows, but the new insights provided by the new approach are crucial.

Some examples.

• One of the important classical results is the Riemann-Roch theorem for a projective curve C of genus g; this theorem gives the dimension of the vector space of functions on C having at most a prescribed set of poles there. This was extended to projective varieties of dimension two, but a general, higher-dimensional analogue seemed out of reach using older methods. The newer approach provided a natural way of looking at the general problem; today, through the work of Hirzebruch, Grothendieck and others, there exists a neat cohomological generalization to arbitrary dimension. (See [**6**].) Riemann-Roch type results are important in dealing with classification problems.

• An old and venerated problem in algebraic geometry was to construct a nonsingular projective variety having a prescribed function field. Riemann assumed it; in 1945 Zariski, after Herculean efforts, solved the problem over groundfields of characteristic zero for varieties of dimension not more than three.

For a time it seemed that if the acknowledged expert Zariski couldn't resolve singularities in higher dimensions, then they just couldn't be resolved! But the power of the newer methods rekindled new hope and efforts. Finally, in 1964, H. Hironaka solved the problem in characteristic zero for arbitrary dimension [**5**].

• In §8 we saw how "Fermat's Last Theorem" could be looked at as the problem of showing that the curve $x^n + y^n = 1$ in $\mathbb{Q}_{xy}$ has no nontrivial points. One can more generally consider the equation

$$a_0x_0^{n_0} + \cdots + a_rx_r^{n_r} = b \tag{1}$$

and ask for the number of solutions in various fields. Many congruence results and problems in number theory can be looked at as finding the number of such solutions when the field is finite. Just as with Fermat's Last Theorem, this too can be looked at geometrically, this time as a problem of counting the number of points of the variety defined by (1), considered over a finite field.

In a 1949 paper [**12**], Weil formulated some conjectures about the solutions to (1). In dimension 1 ($r = 1$ in (1)), he had answered these questions in [**11**]; various higher-dimensional examples pointed towards their truth for arbitrary $r > 1$, but, as is so often the case in algebraic geometry, generalizing to higher dimensions required new methods. In brief, the conjectures, if true, would establish fundamental connections between the number of solutions to (1) in finite fields (as reflected in a "zeta function" attached to (1)'s variety; such a zeta function is pregnant with number-theoretic information, just as is the usual zeta function familiar in complex variables) and the topology (actually, Betti numbers) of projective varieties over $\mathbb{C}$. These conjectures stimulated a great amount of work, especially to find a cohomology for varieties over finite fields strong enough to yield the conjectures. Important contributions were made by Lang-Weil [**8**], Dwork [**2**], Grothendieck ([**3**], for instance), Deligne and others. All the conjectures have now been settled in the positive; the most resistant part was finally cracked by Deligne in 1973 [**1**]. At the same time, this has settled various conjectures in number theory.

The above examples give a little idea of some of the accomplishments achieved using these newer methods. In addition to yielding concrete results, these methods bring with them a new perspective—a perspective that on one hand has extended algebraic geometry, and on the other hand, has helped unify it.

I close on this thought: often, when a new set of ideas is introduced into mathematics, it all initially seems like a tremendous blizzard, with piles and piles of snow (a charge surely not escaped by algebraic geometry). But after a time it melts down to its essence—a few big ideas and

major theorems. The new ideas seem after a while to be simple and natural, and future generations build on them to arrive at mathematics that is deeper, more general and more unified. Yet there will always be concrete examples like curves, surfaces and numbers! They will never go out of style, and will always serve as a source of and testing ground for new directions and new ideas.

References

1. P. Deligne, La conjecture de Weil, I, Publ. Math. IHES, 43 (1974) 273–307.
2. B. Dwork, On the rationality of the zeta function of an algebraic variety, Amer. J. Math., 82 (1960) 631–648.
3. A. Grothendieck, Formule de Lefschetz et rationalité des fonctions L, Séminaire Bourbaki 279 (1965).
4. R. Hartshorne, Algebraic Geometry, Springer-Verlag, New York, 1977.
5. H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. of Math., 79 (1964) 109–326.
6. F. Hirzebruch, Topological Methods in Algebraic Geometry, Springer-Verlag, Heidelberg, 1966.
7. K. Kendig, Elementary Algebraic Geometry, Springer-Verlag, New York, 1977.
8. S. Lang and A. Weil, Number of points of varieties in finite fields, Amer. J. Math., 76 (1954) 819–827.
9. R. Walker, Algebraic Curves, Dover Publications, New York, 1962.
10. A. Weil, Foundations of Algebraic Geometry, 2nd ed., American Mathematical Society, Providence, RI, 1962.
11. ______, Sur les Courbes Algébriques et les Variétés qui s'en Déduisent, Hermann, Paris, 1948.
12. ______, Number of solutions of equations over finite fields, Bull. Amer. Math. Soc., 55 (1949) 497–508.

STRUCTURING MATHEMATICAL PROOFS

URI LERON
School of Education, University of Haifa, Haifa, Israel

1. Introduction. Mathematical proofs are normally presented in a step-by-step, "linear" fashion, proceeding unidirectionally from hypotheses to conclusion. While this age-old and venerable method may be well suited for securing the validity of proofs, it is nonetheless unsuitable for a second, highly important role of most presentations—that of mathematical *communication*.

In this article an alternative method, called the "structural method," is proposed. The method, triggered by recent ideas from computer science, is intended to increase the comprehensibility of mathematical presentations while retaining their rigor. The basic idea underlying the structural method is to arrange the proof in *levels*, proceeding from the top down; the levels themselves consist of short autonomous "modules," each embodying one major idea of the proof.

The top level gives in very general (but precise) terms the main line of the proof. The second level elaborates on the generalities of the top level, supplying proofs for unsubstantiated statements, details for general descriptions, specific constructions for objects whose existence has been merely asserted, and so on. If some such subprocedure is itself complicated, we may choose to give it in the second level only a "top-level description," pushing the details further down to

---

Uri Leron received his Ph.D. in 1972 from the Hebrew University of Jerusalem under S. A. Amitsur. In the years 1972–1980 he did research in ring theory and taught at the departments of mathematics of the University of Oregon, U.C.L.A., and the Technion—Israel Institute of Technology. Since 1980 he has been a Senior Lecturer at the School of Education, Haifa University. He is particularly interested in ways of communicating mathematical ideas that are not apparent from the mathematical formalism. He is also interested in the potential of computer programming for developing mathematical thinking. In connection with the latter he is currently carrying out an experimental project in which sixth graders learn programming through Turtle Graphics in the Logo language.

major theorems. The new ideas seem after a while to be simple and natural, and future generations build on them to arrive at mathematics that is deeper, more general and more unified. Yet there will always be concrete examples like curves, surfaces and numbers! They will never go out of style, and will always serve as a source of and testing ground for new directions and new ideas.

References

1. P. Deligne, La conjecture de Weil, I, Publ. Math. IHES, 43 (1974) 273–307.
2. B. Dwork, On the rationality of the zeta function of an algebraic variety, Amer. J. Math., 82 (1960) 631–648.
3. A. Grothendieck, Formule de Lefschetz et rationalité des fonctions L, Séminaire Bourbaki 279 (1965).
4. R. Hartshorne, Algebraic Geometry, Springer-Verlag, New York, 1977.
5. H. Hironaka, Resolution of singularities of an algebraic variety over a field of characteristic zero, Ann. of Math., 79 (1964) 109–326.
6. F. Hirzebruch, Topological Methods in Algebraic Geometry, Springer-Verlag, Heidelberg, 1966.
7. K. Kendig, Elementary Algebraic Geometry, Springer-Verlag, New York, 1977.
8. S. Lang and A. Weil, Number of points of varieties in finite fields, Amer. J. Math., 76 (1954) 819–827.
9. R. Walker, Algebraic Curves, Dover Publications, New York, 1962.
10. A. Weil, Foundations of Algebraic Geometry, 2nd ed., American Mathematical Society, Providence, RI, 1962.
11. ______, Sur les Courbes Algébriques et les Variétés qui s'en Déduisent, Hermann, Paris, 1948.
12. ______, Number of solutions of equations over finite fields, Bull. Amer. Math. Soc., 55 (1949) 497–508.

STRUCTURING MATHEMATICAL PROOFS

URI LERON
School of Education, University of Haifa, Haifa, Israel

1. Introduction. Mathematical proofs are normally presented in a step-by-step, "linear" fashion, proceeding unidirectionally from hypotheses to conclusion. While this age-old and venerable method may be well suited for securing the validity of proofs, it is nonetheless unsuitable for a second, highly important role of most presentations—that of mathematical *communication*.

In this article an alternative method, called the "structural method," is proposed. The method, triggered by recent ideas from computer science, is intended to increase the comprehensibility of mathematical presentations while retaining their rigor. The basic idea underlying the structural method is to arrange the proof in *levels*, proceeding from the top down; the levels themselves consist of short autonomous "modules," each embodying one major idea of the proof.

The top level gives in very general (but precise) terms the main line of the proof. The second level elaborates on the generalities of the top level, supplying proofs for unsubstantiated statements, details for general descriptions, specific constructions for objects whose existence has been merely asserted, and so on. If some such subprocedure is itself complicated, we may choose to give it in the second level only a "top-level description," pushing the details further down to

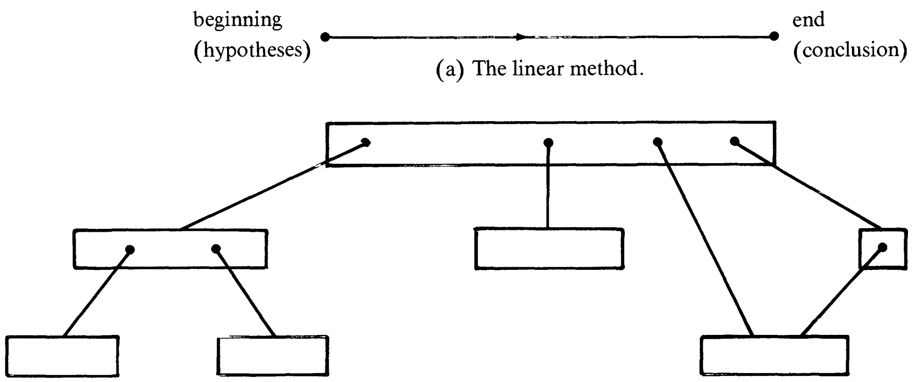
---

Uri Leron received his Ph.D. in 1972 from the Hebrew University of Jerusalem under S. A. Amitsur. In the years 1972–1980 he did research in ring theory and taught at the departments of mathematics of the University of Oregon, U.C.L.A., and the Technion—Israel Institute of Technology. Since 1980 he has been a Senior Lecturer at the School of Education, Haifa University. He is particularly interested in ways of communicating mathematical ideas that are not apparent from the mathematical formalism. He is also interested in the potential of computer programming for developing mathematical thinking. In connection with the latter he is currently carrying out an experimental project in which sixth graders learn programming through Turtle Graphics in the Logo language.

lower levels. And so we continue down the hierarchy of subprocedures, each supplying more details to plug in holes in higher levels, until we reach the bottom where (to borrow W. W. Sawyer's metaphor [**10**, p. 222]) all the leaks are plugged and the proof is watertight.

The top level is normally very short and free of technical (i.e., notational, computational, etc.) details. Thus it can be grasped at one glance, yielding an overview of the proof. (Note that the very term "overview" suggests view from the top.) The bottom level is quite detailed, resembling in this respect the standard linear proof. However, these details now appear only after their role in the proof is determined. Furthermore, they are now organized into conceptual units (the modules), each clearly and explicitly connected to its appropriate place in the total hierarchy. The intermediate levels facilitate a smooth transition from the generalities of the top level to the details of the bottom, from the global to the local perspective.

The two approaches are compared pictorially in Fig. 1. The linear method is represented by an oriented line segment (a), the structural method by a "structure diagram" (b). The structure diagram displays the levels, the modules and their interconnections. In each box, or module, the argument flows linearly, but it is very short and "flat" (no complex nesting patterns of sublevels); thus, again, it can be grasped at a glance. Incidentally, the description above applies not only to proofs, but to other mathematical procedures such as definitions, constructions, algorithms and examples as well.



(a) The linear method.

(b) The structural method.

FIG. 1. The two methods of presentation.

One may think of the structural approach as viewing the proof (which is at ground level) from a tall building. When viewing from the top we see the whole proof at a glance, but only in vague outline—no details can be discerned. As we descend the levels of the building, a zooming effect occurs: our view encompasses smaller and smaller segments of the proof, but these are seen with more and more clarity.

I anticipate a few objections some readers may have at this point. The first two objections, if carried to extremes, are, respectively, that what I am saying is wrong and that what I am saying is trivially true. More specifically, the first one may be expressed by the statement that "there is nothing wrong with the linear method; after all it was good enough for me and for generations of mathematicians." My answer is that we must look for insights in our students' behavior, not merely in ourselves. After all, most of us are successful survivors of the standard method, so we make up an extremely biased sample to draw conclusions from. (For a fierce attack on the linear method cf. [**6**, pp. 115–119].)

The second objection is represented by the following statement: "But your structural method is

what I've been doing in my classes all along." This statement, which has been expressed by some acknowledged excellent teachers, actually *supports* my argument. If the structural method provides a coherent and explicit system of presentation, whereby many of the "options" of good teachers become standard—it will have achieved its purpose. Note again that these options (e.g., a short overview of a long and complicated proof) not only become standard; they actually become part of the "official," formal proof.

A third argument that has been raised actually grants that knowledge of the structure of a proof is essential for its understanding. However (it goes on), this kind of knowledge is best left to the students to discover for themselves. My response to this subtler variant of the "it was good enough for me" argument is that indeed students should discover the structure for themselves; only they don't! And experience shows that this kind of task is beyond the capabilities of most undergraduates with standard mathematical training. In fact it may turn out to be an important teaching strategy within the structural method, to explicitly train students to structure linear proofs.

Finally, it should be stressed that the proposed modification bears directly only on one part of the teaching-learning process: the formal presentation of mathematical procedures as they commonly occur in lectures, textbooks and journal articles. By no means is it meant to replace the informal and artful devices that good expositors have always been employing to enhance the full and active participation of the student in this process, such as intuition, heuristics, personal metaphors, humor and even acting. After all, if the student is not listening, it matters little what we are saying!

2. Examples. I do not know of any way to prove (or disprove) my claims on the merits of the structural method. So in order to judge these claims we must rely on the powerful yet imprecise and subjective tools of our experience, intuition, reflections and observations; and a good way of activating these tools here is to carry out some mathematical "case studies."

On one level, these case studies are simply examples: they help us clarify and make more concrete the general concepts discussed previously. On a different level, they can function as a kind of mental experiment. That is, one can pretend to be a student reading these proofs for the first time (or an instructor teaching them to a class), and thus recapture mentally some of the real experience of learning (or teaching) by these methods.

Ideally we should study some long and complicated proofs, where the structure is buried deeply in the linear presentation (e.g., Brouwer's Fixed-Point Theorem), but we are limited here to theorems whose proofs are relatively short and elementary. Our examples will thus not be paradigmatic in the sense of exhibiting all the desired properties at once; rather, each will illuminate some of the issues only.

2.1. *The Infinite Number of Primes.* Euclid's original proof that there are infinitely many primes is so short and clear that it can be considered a one-level proof. So let us move further.

The odd prime numbers fall naturally into two classes: those leaving a remainder of 1 upon division by 4, and those leaving a remainder of 3. I shall call them "monadic" and "triadic" for short. Thus a number is monadic if it can be represented as $4k + 1$, and triadic if it can be represented as $4l + 3$ for some integers k and l.

It is natural to ask how the prime numbers are distributed between these two classes, and it turns out that there are infinitely many in *each* of them. (These are special cases of the following theorem of Dirichlet: *In any arithmetic progression* $\{an + b|n = 1,2,3,\dots\}$, *where a and b are relatively prime, there are infinitely many primes.*) We shall consider here only the simpler case, that of the triadic numbers.

THEOREM. *There exist infinitely many triadic primes* (*i.e., numbers of the form* $4k + 3$).

The first proof is patterned after a popular and well-known book.

Proof in the linear style. Consider a product of two monadic numbers:

$$(4k+1)(4l+1) = 4k \cdot 4l + 4k + 4l + 1 = 4(4kl + k + l) + 1,$$

which is again monadic. Similarly, the product of any number of monadic numbers is monadic.

Now assume the theorem is false, so there are only finitely many triadic primes, say $p_1, p_2, \ldots, p_n$. Define $M = 4p_1p_2 \cdots p_n - 1$.

If $p_i | M$, then $p_i | 1$ since $p_i | 4p_1p_2 \cdots p_n$. Since this is impossible, we conclude that no p_i divides M. Also 2 does not divide M as M is odd. Thus all M's prime factors are monadic, hence M itself must be monadic. But $M = 4p_1p_2 \cdots p_n - 1 = 4(p_1p_2 \cdots p_n - 1) + 3$ is clearly triadic—a contradiction. Thus the theorem is proved.

REMARK 1. Note that the student is led blindly and passively through the sequence of steps, which he or she must follow without ever being told the general plan of the proof and for what purpose the various steps are taken. Why, for example, should we "consider a product of two monadic numbers?" Why should we define a number M and why this particular way? Why should we check whether p_i and 2 divide M or not?

These questions should naturally arise in the mind of the alert student as he is going through the various steps of the proof, but the answers are not apparent in the presentation. And even the persistent student who does find the answers can do so only at the end of the proof. Worse yet, for many students, understanding (or learning) a proof has come to mean merely checking the validity of the deduction in each step. Thus they have stopped even asking these questions, let alone finding the answers.

REMARK 2. I hope you have read the proof with the student's eyes, not yours, since to a mature mathematician this proof is still very easy. However, similar frustration with the linear method can be experienced by mature mathematicians too, e.g., when reading a 15-page, strictly linear proof in a research journal.

Proof in the Structural Style.

Level 1. Suppose the theorem is false and let $p_1, p_2, \ldots, p_n$ be all the triadic primes. We construct (in Level 2) a number M having the following two properties:

(a) M as well as all its factors are different from $p_1, p_2, \ldots, p_n$;

(b) M has a triadic prime factor.

These two properties clearly produce a contradiction, as we get a triadic prime which is not one of $p_1, p_2, \ldots, p_n$. Thus the theorem is proved.

In the Elevator. (The elevator, as a metaphor for the intermediate process of descending in levels, offers a convenient place to discuss heuristic and other informal issues concerning the next level.)

How shall we approach the definition of M? In light of Euclid's classical proof, it is natural to try $M = p_1p_2 \cdots p_n + 1$. This indeed meets requirement (a), but not (b). In fact, since for all we know M itself may turn out to be prime, it must be triadic to meet (b).

Thus a natural second guess is $M = 4p_1p_2 \cdots p_n + 3$. However, this has another "bug": since one of the p_i's is 3, M is divisible by 3, in violation of (a). But this bug, once discovered, is easy to fix: simply eliminate 3 from the product in the bugged definition.

Level 2. Let $M = 4p_2 \cdots p_n + 3$ (we assume $p_1 = 3$). We show that M satisfies the two requirements from Level 1.

Requirement (a) means that no p_i should divide M. Indeed, $p_2, \ldots, p_n$ do not divide M as they leave a remainder of 3; and 3 does not divide M as it does not divide $4p_2 \cdots p_n$.

As for requirement (b), suppose on the contrary that all of M's prime factors were monadic. Then M, as a product of monadic numbers, would itself be monadic (Lemma, Level 3)—a contradiction. Thus (a) and (b) are satisfied.

Level 3: LEMMA. *A product of monadic numbers is again a monadic number.*

(The proof is as given above.)

REMARK 3. Level 1 demonstrates two important features of the top level. First, it can be grasped at one glance, and second, it gives the essence of the proof. Consider for example the introduction of M. What is stressed is its role in the proof, the properties it should enjoy, and how these are used to achieve the goal. The actual, detailed construction, as well as the proof that M satisfies the required properties, are pushed down to a lower level. Thus the top level indeed gives a global view of the proof.

REMARK 4. The "elevator" trick is a side benefit of the structural method, not an "official" part of it. Authors and instructors often like to inject informal remarks (mainly of a heuristic character) into the ongoing formal proof. In the linear method, where the argument is presented as an unbroken sequence of steps, there is no natural place to do it. So authors have been using various display tricks such as footnotes, different font, brackets and so on, to distinguish these remarks from the formal proof. In lectures, students are often confused as to what constitutes the formal proof and what are the auxiliary remarks.

In the structural approach, since the proof is divided anyhow into independent, relatively short modules, it is always possible to collect all the informal remarks and present them during the "elevator trip," that is, while descending to the next level.

REMARK 5. The successive debugging in the search for M illustrates an important point. I often ask my students to keep a protocol of the *process* of finding a proof, besides the final product. From these protocols it appears that no one reaches the correct M without going through some debugging, roughly of the kind given here. Since this is an essential part of the creative process of *doing* mathematics, I believe it should be made explicit and discussed with the students.

REMARK 6. Compare the location of the lemma on the product of monadic numbers in the two approaches. In the linear approach it appears in the beginning of the proof (or sometimes as a separate lemma *before* the main proof). The reader can have no idea why it is needed, but has to follow its proof nonetheless, and then keep the unused lemma in memory until it is needed at the *end* of the proof. In the structural approach the lemma is only announced when the need for it arises in the main proof. The lemma is used to complete the main proof and only then, in a lower level, its proof is given.

REMARK 7. The case of the monadic primes is more involved and will not be discussed here. However, it should be pointed out that in the structural method we have at least a clear beginning; in fact, the top level of the proof remains exactly the same, except for the exchange "triadic" $\leftrightarrow$ "monadic." (One says that the two proofs exhibit a "top-level similarity.") The difficulty, as well as the "meat" of the proof, lie in the lower levels: the number M, promised in the top level, is not easy to deliver.

2.2. *A Theorem on Limits*

THEOREM. *If* $\lim_{x\to a} f(x) = L$ *and* $\lim_{x\to a} g(x) = M$, *then* $\lim_{x\to a} f(x)g(x) = LM$.

The following proof is not a caricature of the linear style; it is written strictly according to the "official" principles of that style and, in fact, is taken from a real calculus textbook.

Proof in the Linear Style. Let $\epsilon > 0$ be given, and let η be the smaller of $\sqrt{\epsilon/3}$ and $\epsilon/3(1 + |L| + |M|)$. Since $\lim_{x\to a} f(x) = L$, there exists a $\delta_1 > 0$ such that $|f(x) - L| < \eta$ whenever $0 < |x - a| < \delta_1$. Similarly, there exists a $\delta_2 > 0$ such that $|g(x) - M| < \eta$ whenever $0 < |x - a| < \delta_2$. Let δ be the smaller of δ_1 and δ_2. Now if $0 < |x - a| < \delta$, then $0 < |x - a| < \delta_i$, $i = 1, 2$, and so we have:

$$\begin{aligned}|f(x)g(x)-LM| &= |L(g(x)-M)+M(f(x)-L)+(f(x)-L)(g(x)-M)|\\ &\leqslant |L||g(x)-M|+|M||f(x)-L|+|f(x)-L||g(x)-M|\\ &< |L|\epsilon/3(1+|L|+|M|)+|M|\epsilon/3(1+|L|+|M|)+\sqrt{\epsilon/3}\cdot\sqrt{\epsilon/3}\\ &\leqslant \epsilon/3+\epsilon/3+\epsilon/3=\epsilon. \qquad \text{q.e.d.}\end{aligned}$$

Fortunately, many instructors know better. They let the student in on the secret of how these mysterious quantities η and δ are actually *discovered.* But in so doing the direction of the argument is reversed, and eventually they have to abandon this unorthodox discussion and recast the official proof in more-or-less the form above (or at least mention that this recasting *should* be done).

The argument in the following structured proof resembles this informal discussion but at the same time it is quite formal and rigorous. Thus the structural approach brings closer the human process and the formal-deductive one.

A Structured Proof.

Level 1. Let $\epsilon > 0$ be given. We find (in level 2) a $\delta > 0$ such that $0 < |f(x)g(x) - LM| < \epsilon$ whenever $0 < |x - a| < \delta$. Thus the theorem is proved.

In the Elevator. We have to show that the expression $|f(x)g(x) - LM|$ can be made as small as we please. To this end, we try to bound it by an expression we *know* can be made small. Such expressions are $|f(x) - L|, |gx) - M|$ and multiples of these by a constant and by each other. After some trial and error the following expression emerges:

$$(*)\quad f(x)g(x)-LM = L(g(x)-M)+M(f(x)-L)+(f(x)-L)(g(x)-M).$$

Level 2. Using the equality $(*)$ we have

$$\begin{aligned}|f(x)g(x)-LM| &= |L(g(x)-M)+M(f(x)-L)+(f(x)-L)(g(x)-M)|\\ &\leqslant |L||g(x)-M|+|M||f(x)-L|+|f(x)-L||g(x)-M|.\end{aligned}$$

We find a $\delta > 0$ (in Level 3) such that whenever $0 < |x - a| < \delta$, each of the terms on the right-hand side is smaller than $\epsilon/3$. Thus the left-hand side is smaller than ϵ, as required.

In the Elevator. To get $|L||g(x) - M| < \epsilon/3$, we try to make $|g(x) - M| < \epsilon/3|L|$. However, there is a bug here: this only works if $L \neq 0$. One way of correcting this bug is to replace $|L|$ by $1 + |L|$. The case of $|M||f(x) - L|$ is similar. Finally, to get $|f(x) - L||g(x) - M| < \epsilon/3$, we make each of the factors smaller than $\sqrt{\epsilon/3}$.

Level 3. We choose positive $\delta_1, \delta_2, \delta_3, \delta_4$ such that the following hold:

$$\begin{aligned}&|f(x)-L| < \epsilon/3(1+|M|) &&\text{whenever} && 0<|x-a|<\delta_1;\\ &|g(x)-M| < \epsilon/3(1+|L|) &&\text{whenever} && 0<|x-a|<\delta_2;\\ &|f(x)-L| < \sqrt{\epsilon/3} &&\text{whenever} && 0<|x-a|<\delta_3;\\ &|g(x)-M| < \sqrt{\epsilon/3} &&\text{whenever} && 0<|x-a|<\delta_4.\end{aligned}$$

(Such δ_i's exist since L and M are the limits of $f(x)$ and $g(x)$ respectively.) Now let δ be the smallest of $\delta_1, \delta_2, \delta_3, \delta_4$, so that if $0 < |x - a| < \delta$ then $0 < |x - a| < \delta_i$, $i = 1, 2, 3, 4$. Then whenever $0 < |x - a| < \delta$, the expressions $|L||g(x) - M|$, $|M||f(x) - L|$ and $|f(x) - L||g(x) - M|$ all become smaller than $\epsilon/3$. Thus δ satisfies the requirements of Level 2.

REMARK. As seen from this example, structural proofs are longer to deliver, but (I believe) shorter to digest. In fact, they are longer because they contain more information (namely, the structure of the proof), and it is this very information which makes them more learnable,

illuminating and human. Thus in switching to structured proofs we simply agree to share with our students (or readers, etc.) more of what we know about the proof. And it is my belief that the loss in economy is more than balanced by the gain in learning.

2.3. *A High-School Level Problem...*

We next take up the following ruler-and-compass construction problem:

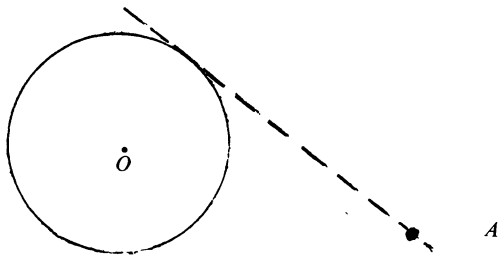


FIG. 2. A construction problem.

Construct a tangent to a given circle from a given point outside the circle.

The linear description of the construction is as follows (refer to Fig. 3).

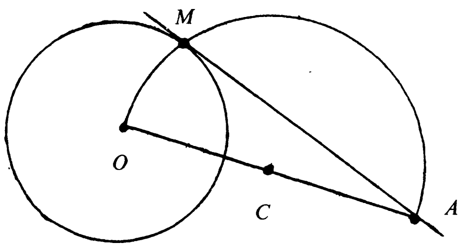


FIG. 3. The construction.

Draw the line segment OA and find its midpoint C. Around C as center draw a circle with radius $|OC|$. Let M be one point of intersection of the two circles. Then AM is the required tangent.

In general, the solution of construction problems in school usually proceeds in three stages. First, an analysis of the problem is carried out and the construction is discovered; second, the construction is described formally (in reverse order!); third, a proof is given that the construction indeed meets all the requirements.

In the structural presentation, all three stages are combined in one coherent process, thus again uniting the human (the first stage) with the mathematical (the second and third).

A Structured Proof.

Level 1. We find (in Level 2) a point M on the circle so that the line AM is perpendicular to the radius OM. Then (by a well-known theorem) AM is a tangent, as required.

In the Elevator. How shall we find the point M? We follow a well-known heuristic principle ([**9**, Ch. 1]): Look at the two properties determining M and construct the two corresponding loci; then construct M as the intersection of these two loci.

Level 2. Let M be a point of intersection of the given circle with the locus of points where the segment OA subtends a right angle. That is, M is a point of intersection of the given circle and the circle (constructed in Level 3) having OA as diameter.

Level 3. To construct the circle with segment OA as diameter, let C be the midpoint of segment OA, then draw the circle with C as center and $|OC|$ as radius. Clearly OA is a diameter in this circle, as required.

2.4. ... *And a More Advanced One*. The last example represented what seems, more or less, the lowest level of complexity in which arranging the proof in levels is still advantageous. While there is no corresponding upper limit (the more complex the proof, the more helpful is a structural presentation), we are nonetheless limited in our choice of case studies by the scope of this paper.

Pushing towards this limit, we now take up a theorem of central importance in linear algebra, namely the theorem on canonical forms for Hermitian operators and matrices. The theorem has several variants (the best known of which being, perhaps, the one on orthogonal diagonalization of symmetric matrices), and is a close relative of the Spectral Theorem. To save space, we state it directly in its more technical, ready-to-prove form.

THEOREM. *Let V be a finite-dimensional complex inner-product space, and let T be a Hermitian operator on V. Then V has an orthonormal basis of characteristic vectors of T.*

(Recall that T is Hermitian if $\langle Tu, v\rangle = \langle u, Tv\rangle$ for all vectors u and v in V, where $\langle x, y\rangle$ denotes the inner product of x and y.)

The following proof is self-contained modulo the standard material on operators and inner-product spaces; that is, it assumes no knowledge of Hermitian operators beyond the definition. See, e.g., [**2**], [**3**], or [**4**] for the standard facts, as well as for standard proofs of the theorem. Usually, the contents of this proof are spread out over several lemmas and theorems, thus achieving a measure of structuring even in the standard presentations. The structuring, however, is incomplete and the presentation proceeds *bottom-up*; that is, the auxiliary results (lower levels) appear first and the main argument (top level) last. (Note, however, the top-down commentary in [**3**, p. 301].)

A Structured Proof.

Level 1. Let U be the subspace of V spanned by all the characteristic vectors of T. We prove two assertions about U:

(a) $U = V$, i.e., the characteristic vectors of T span the whole space (Level 2.1);

(b) U has an orthonormal basis of characteristic vectors of T (Level 2.2).

Clearly, these two assertions yield the conclusion of the theorem.

REMARK. From here on the proof branches to two independent subproofs, rooted at 2.1 and 2.2. Note that while the presentation proceeds strictly top-down, it does not force a top-down *reading* of the proof. Some readers may prefer, for example, to read through the first branch (2.1) all the way to the bottom, then return to Level 2 and start on the second branch (2.2).

In what follows we have added a number (in parentheses) to the title of each module. This number is a back-reference to the "calling" module, i.e., the one where the present module is referenced. A more compact representation of the interconnections within the hierarchy is given in the structure diagram (FIG. 4).

Level 2.1 (1). To prove $U = V$ we prove the equivalent statement $U^{\perp} = \{0\}$, where $U^{\perp}$ is the orthogonal complement of U. This in turn will follow from the following two assertions:

(c) $U^{\perp}$ is T-invariant (Level 3.1).

(d) Every nonzero T-invariant subspace of V contains a characteristic vector of T (Level 3.2).

Since $U^{\perp}$ cannot contain a characteristic vector of T (this would contradict $U \cap U^{\perp} = \{0\}$), we must have $U^{\perp} = \{0\}$, hence $U = V$.

Level 2.2 (1). Let $\lambda_1, \dots, \lambda_k$ be the distinct characteristic values of T, and let $U_1, \dots, U_k$ be their respective characteristic spaces:

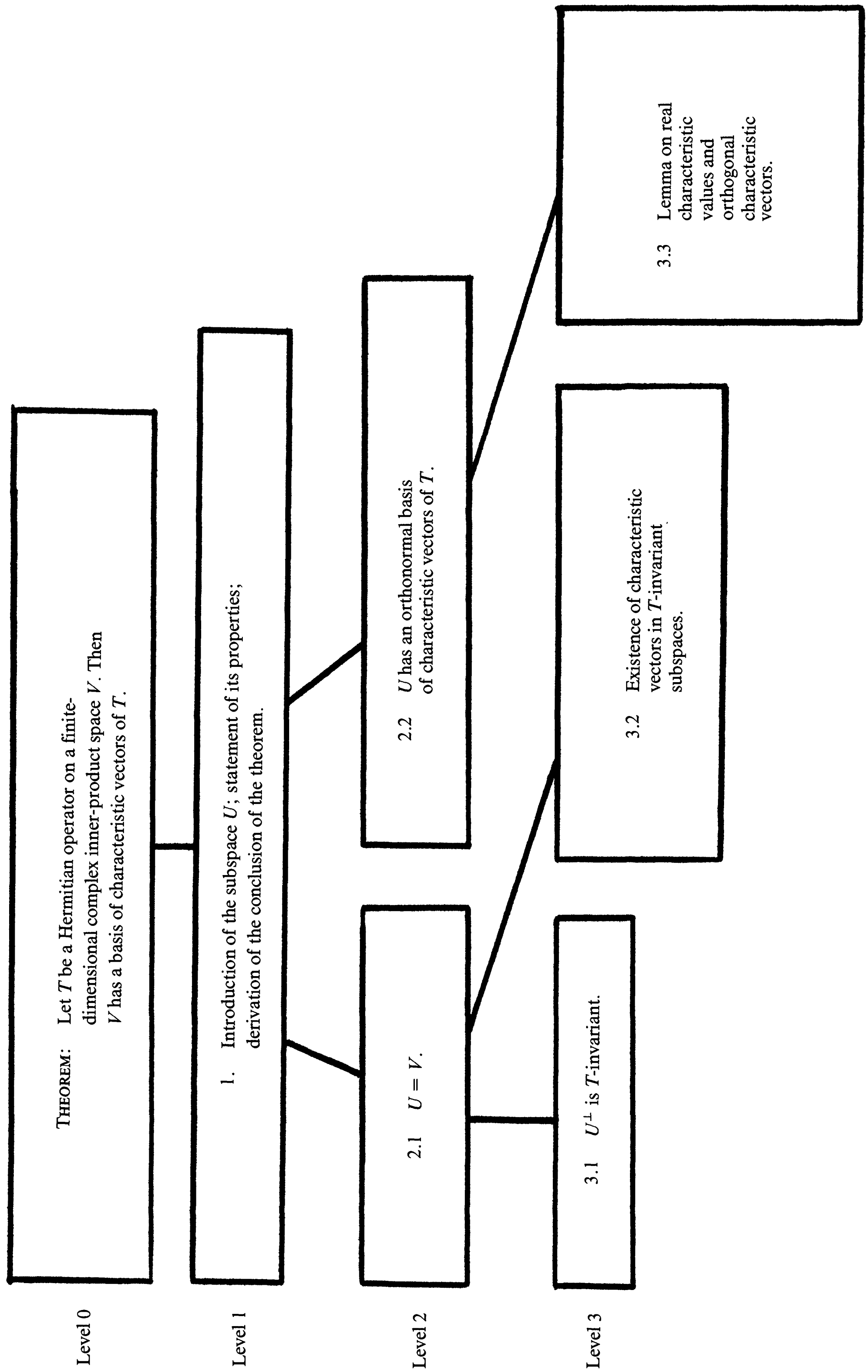


FIG. 4. Structure of the proof of the Theorem.

$$U_i = \{v \in V | Tv = \lambda_i v\}.$$

By the Gram-Schmidt orthogonalization process we can construct an orthonormal basis B_i to each subspace U_i, and it turns out that without any more care on our part, the set $B = B_1 \cup \cdots \cup B_k$ already forms an orthonormal basis for V. This is due to the following important lemma (proved in Level 3.3):

Characteristic vectors of T that belong to distinct characteristic values are orthogonal.

Thus if $i \neq j$, B_i and B_j are orthogonal, hence B is indeed an orthonormal basis for U. Since B consists of characteristic vectors of T, it satisfies all the requirements of Level 1 (b).

Level 3.1 (2.1). $U^{\perp}$ is T-invariant.

Level 3.2 (2.1). If W is a nonzero T-invariant subspace of V, then W contains a characteristic vector of T.

Level 3.3 (2.2). LEMMA. Let T be a Hermitian operator. Then

(a) *All the (complex) characteristic values of T are real.*

(b) *Characteristic vectors of T that belong to distinct characteristic values are orthogonal.*

We leave out the proofs of 3.1, 3.2 and 3.3, since Level 3 is the bottom level and the proofs appear in it in their standard (linear) form (see, e.g., [**4**, pp. 312–313]). Note, however, that it is only in 3.1 and 3.3 that the main hypothesis of the Theorem (namely that T is Hermitian) finally enters the proof. In contrast, 3.2 is true for operators on complex spaces in general.

REMARK. There is a similar theorem for normal operators, and the similarity is 3-level deep, counting the theorem itself as level 0 of the proof. Since 3.2 is true for operators on complex spaces in general, only 3.1 and 3.3 need to be adjusted.

We conclude this subsection with a structure diagram of the proof (Fig. 4), displaying the hierarchy of modules and their interconnections.

3. Summary: Benefits of the Structural Method.

3.1 *More Communicative Proofs.* The main benefit of presentations in the structural style is, hopefully, that ideas behind the formal proofs are better communicated. As we have seen, the main idea is given in the top level (or two), auxiliary ideas are packaged in autonomous modules, and the interconnections between the separate ideas are made explicit through the structure diagram.

By analyzing the foregoing examples, we can be a bit more specific about what is meant by the "main idea" of a proof, and the different treatment given to it by the two methods. The main idea often lies in the construction of a new, intermediate object, *the pivot*, to mediate between the hypotheses and the conclusion. (In our examples the pivots are the number M in 2.1, the numbers η, δ, δ_i in 2.2, the point M in 2.3, and the orthonormal basis of characteristic vectors, B, in 2.4.) Since the pivot occupies a central position in the proof and is directly connected to all its parts, it offers a vantage point from which to view the global architecture of the proof; and precisely this view is given in the top level of the structured proof (cf. the examples). Here the pivot is introduced by a statement of its essential properties and is immediately used to derive the conclusion of the theorem. The detailed definition, as well as proof of the postulated properties and questions of existence, are all pushed down to lower levels.

In the linear approach the pivot is treated poorly (from the learner's point of view) and its potential for revealing the architecture of the proof is wasted. Rather to the contrary, it is here where the proof most resembles pulling a rabbit from the hat (cf., e.g., the introduction of the number η in the linear proof of 2.2 or, for an extreme example, the element $f(a, b, c)$ in [**5**, p. 274]). The pivot is usually introduced near the beginning of the proof by a bare statement of its definition, which often appears extremely bizarre and complicated. Such definitions have an intimidating, even paralyzing, effect on many students when introduced too abruptly. The words

of Courant and Robbins, taken from a slightly different context ([**1**, p. 292]), are appropriate here: "There is an unfortunate, almost snobbish attitude on the part of some writers of textbooks, who present the reader with this definition without a thorough preparation, as though an explanation were beneath the dignity of a mathematician." Note, in contrast, the "untricking" effect a structured presentation has on the definition of the pivot: the actual detailed definition is given only after its role in the proof and the reasons for its particular form are made clear in the upper levels.

It should be remarked that the term "main idea" is used here more in the sense of an abstract or overview, and in no way is it meant to imply that mathematically it is the most important idea in the proof. For example, in the proof that there are infinitely many monadic primes (see Remark 7 of 2.1) the main *mathematical* idea may well be to take $M = (2p_1p_2 \cdots p_n)^2 + 1$, which is relatively a low-level idea. Still, I would maintain that awareness of the main idea in the above sense is important for an insightful learning, and is more likely to lead the student away from learning proofs and definitions by rote.

3.2. *Learning Activities.*

Several new, structure-related activities are available to the learner, or the instructor, in view of the structural method. Here is a sample:

—Given the higher levels of a proof, complete the lower levels.

—Take any proof from a standard textbook and find its structure (i.e., arrange the proof in levels). This is not easy, but it is highly valuable and rewarding; and it results in a much deeper understanding of the proof. (For some examples of this activity, see Section 2.)

—This is related to the previous one: Given a 10-page proof, describe it in one page.

—Given two theorems that show some similarity, determine how deep this similarity is; that is, to how many levels of the proof does the similarity extend, counting from the top.

All these activities can be used as self-learning activities, or they can be assigned as homework or test. Not only do they penetrate deeply into the particular subject matter, but they also encourage the learner to reflect on the process of theorem-proving itself.

An additional benefit of the structural method is that of *partial proofs*. The situation often arises, where presenting or studying a proof to the last detail is impossible or undesirable, due to insufficient level of mathematical preparation or motivation on the part of the audience, or simply due to time limitations. Consider, for example, the following cases: survey lectures or articles, mathematics courses for engineering students, an instructor scanning a pile of textbooks in preparation for a course, a mathematician scanning a pile of research articles. In all such cases the structural method allows one to conveniently choose a suitable level of detail for the particular situation, by simply ignoring (perhaps temporarily) some of the lower levels of the proof. This bears some similarity to Pólya's advice concerning the use of "incomplete proofs" ([**8**, "Why proofs?", pp. 219–221]), but note that our partial proofs can always be refined to a complete proof.

3.3 *Conclusion.* Iu. I. Manin has already referred to the human aspect of proofs when he said [**7**, p. 51]: "A good proof is one which makes us wiser." Clearly, then, a good *presentation* of a proof is one which makes the *listener* (or reader) wiser. It is my hope that presentations in the structural style have an improved chance of passing this difficult test.

Acknowledgment. I acknowledge gratefully the helpful comments from many colleagues on an earlier version of this paper.

References

1. R. Courant and H. Robbins, What is Mathematics?, Oxford University Press, New York, 1941.
2. P. R. Halmos, Finite-Dimensional Vector Spaces, 2nd ed., Van Nostrand, Princeton, 1958.
3. I. N. Herstein, Topics in Algebra, Blaisdell, Waltham, Massachusetts, 1964.

4. K. Hoffman and P. Kunze, Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1971.
5. N. Jacobson, Structure of Rings, revised ed., Amer. Math. Soc., Providence, RI, 1964.
6. M. Kline, Why the Professor Can't Teach, St. Martin's Press, New York, 1977.
7. Iu. I. Manin, A Course in Mathematical Logic, Springer-Verlag, 1977.
8. G. Pólya, How to Solve It, 2nd ed., Doubleday, Garden City, NY, 1957.
9. ______, Mathematical Discovery, vol. I, Wiley, New York, 1964.
10. W. W. Sawyer, A Concrete Approach to Abstract Algebra, Freeman, San Francisco, 1959.

WHO GAVE YOU THE EPSILON? CAUCHY AND THE ORIGINS OF RIGOROUS CALCULUS

JUDITH V. GRABINER
424 West 7th Street, Claremont, California 91711

Student: The car has a speed of 50 miles an hour. What does that mean?
Teacher: Given any $\varepsilon > 0$, there exists a δ such that if $|t_2 - t_1| < \delta$, then $\left|\frac{s_2 - s_1}{t_2 - t_1} - 50\right| < \varepsilon$.
Student: How in the world did anybody ever think of such an answer?

* * * * * * * * * * * * *

Perhaps this exchange will remind us that the rigorous basis for the calculus is not at all intuitive—in fact, quite the contrary. The calculus is a subject dealing with speeds and distances, with tangents and areas—not inequalities. When Newton and Leibniz invented the calculus in the late seventeenth century, they did not use delta-epsilon proofs. It took a hundred and fifty years to develop them. This means that it was probably very hard, and it is no wonder that a modern student finds the rigorous basis of the calculus difficult. How, then, did the calculus get a rigorous basis in terms of the algebra of inequalities?

Delta-epsilon proofs are first found in the works of Augustin-Louis Cauchy (1789–1867). This is not always recognized, since Cauchy gave a purely verbal definition of limit, which at first glance does not resemble modern definitions: "When the successively attributed values of the same variable indefinitely approach a fixed value, so that finally they differ from it by as little as desired, the last is called the *limit* of all the others" [**1**]. Cauchy also gave a purely verbal definition of the derivative of $f(x)$ as the limit, when it exists, of the quotient of differences $(f(x + h) - f(x))/h$ when h goes to zero, a statement much like those that had already been made by Newton, Leibniz, d'Alembert, Maclaurin, and Euler. But what is significant is that Cauchy translated such verbal statements into the precise language of inequalities when he needed them in his proofs. For instance, for the derivative [**2**]:

(1) Let δ, ε be two very small numbers; the first is chosen so that for all numerical [i.e., absolute] values of h less than δ, and for any value of x included [in the interval of definition], the ratio $(f(x + h) - f(x))/h$ will always be greater than $f'(x) - \varepsilon$ and less than $f'(x) + \varepsilon$.

Judith V. Grabiner has taught the history of science since 1972 at California State University, Dominguez Hills, where she is Professor of History. After getting a B. S. in Mathematics from the University of Chicago, she received her M. A. and Ph. D. (1966) at Harvard University. She is Book Review Editor of *Historia Mathematica*, Chairman of the Southern California Section of the Mathematical Association of America, and the author of *The Origins of Cauchy's Rigorous Calculus* (M. I. T. Press, 1981). In 1982–1983 she was Visiting Professor of History at U. C. L. A.

This paper is a revised version of a talk given at various mathematics colloquia, at the Summer Meeting of the Mathematical Association of America in Ann Arbor, Michigan, in 1980, and at the New York Academy of Sciences in March, 1982. Some of the research was supported by the National Science Foundation under Grant No. SOC 7907844.

4. K. Hoffman and P. Kunze, Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1971.
5. N. Jacobson, Structure of Rings, revised ed., Amer. Math. Soc., Providence, RI, 1964.
6. M. Kline, Why the Professor Can't Teach, St. Martin's Press, New York, 1977.
7. Iu. I. Manin, A Course in Mathematical Logic, Springer-Verlag, 1977.
8. G. Pólya, How to Solve It, 2nd ed., Doubleday, Garden City, NY, 1957.
9. ______, Mathematical Discovery, vol. I, Wiley, New York, 1964.
10. W. W. Sawyer, A Concrete Approach to Abstract Algebra, Freeman, San Francisco, 1959.

WHO GAVE YOU THE EPSILON? CAUCHY AND THE ORIGINS OF RIGOROUS CALCULUS

JUDITH V. GRABINER
424 West 7th Street, Claremont, California 91711

Student: The car has a speed of 50 miles an hour. What does that mean?

Teacher: Given any $\varepsilon > 0$, there exists a δ such that if $|t_2 - t_1| < \delta$, then $\left|\frac{s_2 - s_1}{t_2 - t_1} - 50\right| < \varepsilon$.

Student: How in the world did anybody ever think of such an answer?

* * * * * * * * * * * * *

Perhaps this exchange will remind us that the rigorous basis for the calculus is not at all intuitive—in fact, quite the contrary. The calculus is a subject dealing with speeds and distances, with tangents and areas—not inequalities. When Newton and Leibniz invented the calculus in the late seventeenth century, they did not use delta-epsilon proofs. It took a hundred and fifty years to develop them. This means that it was probably very hard, and it is no wonder that a modern student finds the rigorous basis of the calculus difficult. How, then, did the calculus get a rigorous basis in terms of the algebra of inequalities?

Delta-epsilon proofs are first found in the works of Augustin-Louis Cauchy (1789–1867). This is not always recognized, since Cauchy gave a purely verbal definition of limit, which at first glance does not resemble modern definitions: "When the successively attributed values of the same variable indefinitely approach a fixed value, so that finally they differ from it by as little as desired, the last is called the *limit* of all the others" [**1**]. Cauchy also gave a purely verbal definition of the derivative of $f(x)$ as the limit, when it exists, of the quotient of differences $(f(x + h) - f(x))/h$ when h goes to zero, a statement much like those that had already been made by Newton, Leibniz, d'Alembert, Maclaurin, and Euler. But what is significant is that Cauchy translated such verbal statements into the precise language of inequalities when he needed them in his proofs. For instance, for the derivative [**2**]:

(1) Let δ, ε be two very small numbers; the first is chosen so that for all numerical [i.e., absolute] values of h less than δ, and for any value of x included [in the interval of definition], the ratio $(f(x + h) - f(x))/h$ will always be greater than $f'(x) - \varepsilon$ and less than $f'(x) + \varepsilon$.

Judith V. Grabiner has taught the history of science since 1972 at California State University, Dominguez Hills, where she is Professor of History. After getting a B. S. in Mathematics from the University of Chicago, she received her M. A. and Ph. D. (1966) at Harvard University. She is Book Review Editor of *Historia Mathematica*, Chairman of the Southern California Section of the Mathematical Association of America, and the author of *The Origins of Cauchy's Rigorous Calculus* (M. I. T. Press, 1981). In 1982–1983 she was Visiting Professor of History at U. C. L. A.

This paper is a revised version of a talk given at various mathematics colloquia, at the Summer Meeting of the Mathematical Association of America in Ann Arbor, Michigan, in 1980, and at the New York Academy of Sciences in March, 1982. Some of the research was supported by the National Science Foundation under Grant No. SOC 7907844.

This one example will be enough to indicate how Cauchy did the calculus, because the question to be answered in the present paper is not, "how is a rigorous delta-epsilon proof constructed?" As Cauchy's intellectual heirs we all know this. The central question is, how and why was Cauchy able to put the calculus on a rigorous basis, when his predecessors were not? The answers to this historical question cannot be found by reflecting on the logical relations between the concepts, but by looking in detail at the past and seeing how the existing state of affairs in fact developed from that past. Thus we will examine the mathematical situation in the seventeenth and eighteenth centuries—the background against which we can appreciate Cauchy's innovation. We will describe the powerful techniques of the calculus of this earlier period and the relatively unimpressive views put forth to justify them. We will then discuss how a sense of urgency about rigorizing analysis gradually developed in the eighteenth century. Most important, we will explain the development of the mathematical techniques necessary for the new rigor from the work of men like Euler, d'Alembert, Poisson, and especially Lagrange. Finally, we will show how these mathematical results, though often developed for purposes far removed from establishing foundations for the calculus, were used by Cauchy in constructing his new rigorous analysis.

The Practice of Analysis: From Newton to Euler. In the late seventeenth century, Newton and Leibniz, almost simultaneously, independently invented the calculus. This invention involved three things. First, they invented the general concepts of differential quotient and integral (these are Leibniz's terms; Newton called the concepts "fluxion" and "fluent"). Second, they devised a notation for these concepts which made the calculus an algorithm: the methods not only worked, but were easy to use. Their notations had great heuristic power, and we still use Leibniz's dy/dx and $\int y\,dx$, and Newton's $\dot{x}$, today. Third, both men realized that the basic processes of finding tangents and areas, that is, differentiating and integrating, are mutually inverse—what we now call the Fundamental Theorem of Calculus.

Once the calculus had been invented, mathematicians possessed an extremely powerful set of methods for solving problems in geometry, in physics, and in pure analysis. But what was the nature of the basic concepts? For Leibniz, the differential quotient was a ratio of infinitesimal differences, and the integral was a sum of infinitesimals. For Newton, the derivative, or fluxion, was described as a rate of change; the integral, or fluent, was its inverse. In fact, throughout the eighteenth century, the integral was generally thought of as the inverse of the differential. One might imagine asking Leibniz exactly what an infinitesimal was, or Newton what a rate of change might be. Newton's answer, the best of the eighteenth century, is instructive. Consider a ratio of finite quantities (in modern notation, $(f(x+h)-f(x))/h)$ as h goes to zero). The ratio eventually becomes what Newton called an "ultimate ratio." Ultimate ratios are "limits to which the ratios of quantities decreasing without limit do always converge, and to which they approach nearer than by any given difference, but never go beyond, nor ever reach until the quantities vanish" [3]. Except for "reaching" the limit when the quantities vanish, we can translate Newton's words into our algebraic language. Newton himself, however, did not do this, nor did most of his followers in the eighteenth century. Moreover, "never go beyond" does not allow a variable to oscillate about its limit. Thus, though Newton's is an intuitively pleasing picture, as it stands it was not and could not be used for proofs about limits. The definition sounds good, but it was not understood or applied in algebraic terms.

But most eighteenth-century mathematicians would object, "Why worry about foundations?" In the eighteenth century, the calculus, intuitively understood and algorithmically executed, was applied to a wide range of problems. For instance, the partial differential equation for vibrating strings was solved; the equations of motion for the solar system were solved; the Laplace transform and the calculus of variations and the gamma function were invented and applied; all of mechanics was worked out in the language of the calculus. These were great achievements on the part of eighteenth-century mathematicians. Who would be greatly concerned about foundations when such important problems could be successfully treated by the calculus? Results were what counted.

This point will be better appreciated by looking at an example which illustrates both the "uncritical" approach to concepts of the eighteenth century and the immense power of eighteenth-century techniques, from the work of the great master of such techniques: Leonhard Euler. The problem is to find the sum of the series

$$1/1 + 1/4 + 1/9 + \cdots + 1/k^2 + \cdots.$$

It clearly has a finite sum since it is bounded above by the series

$$1 + 1/1 \cdot 2 + 1/2 \cdot 3 + 1/3 \cdot 4 + \cdots + 1/(k-1) \cdot k + \cdots,$$

whose sum was known to be 2; Johann Bernoulli had found this sum by treating $1/1 \cdot 2 + 1/2 \cdot 3 + 1/3 \cdot 4 + \cdots$ as the difference between the series $1/1 + 1/2 + 1/3 + \cdots$ and the series $1/2 + 1/3 + 1/4 + \cdots$, and observing that this difference telescopes [**4**].

Euler's summation of $\Sigma_{k=1}^{\infty} 1/k^2$ makes use of a lemma from the theory of equations: given a polynomial equation whose constant term is one, the coefficient of the linear term is the product of the reciprocals of the roots with the signs changed. This result was both discovered and demonstrated by considering the equation $(x-a)(x-b) = 0$, having roots a and b. Multiplying and then dividing out ab, we obtain

$$(1/ab)x^2 - (1/a + 1/b)x + 1 = 0;$$

the result is now obvious, as is the extension to equations of higher degree.

Euler's solution then considers the equation $\sin x = 0$.

Expanding this as an infinite series, Euler obtained

$$x - x^3/3! + x^5/5! - \cdots = 0.$$

Dividing by x yields

$$1 - x^2/3! + x^4/5! - \cdots = 0.$$

Finally, substituting $x^2 = u$ produces

$$1 - u/3! + u^2/5! - \cdots = 0.$$

But Euler thought that power series could be manipulated just like polynomials. Thus, we now have a polynomial equation in u, whose constant term is one. Applying the lemma to it, the coefficient of the linear term with the sign changed is $1/3! = 1/6$. The roots of the equation in u are the roots of $\sin x = 0$ with the substitution $u = x^2$, namely $\pi^2, 4\pi^2, 9\pi^2, \ldots$. Thus the lemma implies

$$1/6 = 1/\pi^2 + 1/4\pi^2 + 1/9\pi^2 + \cdots.$$

Multiplying by π^2 yields the sum of the original series [**5**]:

$$1/1 + 1/4 + 1/9 + \cdots + 1/k^2 + \cdots = \pi^2/6.$$

Though it is easy to criticize eighteenth-century arguments like this for their lack of rigor, it is also unfair. Foundations, precise specifications of the conditions under which such manipulations with infinites or infinitesimals were admissible, were not very important to men like Euler, because without such specifications they made important new discoveries, whose results in cases like this could readily be verified. When the foundations of the calculus were discussed in the eighteenth century, they were treated as secondary. Discussions of foundations appeared in the introductions to books, in popularizations, and in philosophical writings, and were not—as they are now and have been since Cauchy's time—the subject of articles in research-oriented journals.

Thus, where we once had one question to answer, we now have two. The first remains, where do Cauchy's rigorous techniques come from? Second, one must now ask, why rigorize the calculus in the first place? If few mathematicians were very interested in foundations in the eighteenth century [**6**], then when, and why, were attitudes changed?

Of course, to establish rigor, it is necessary—though not sufficient—to think rigor is signifi-

cant. But more important, to establish rigor, it is necessary (though also not sufficient) to have a set of techniques in existence which are suitable for that purpose. In particular, if the calculus is to be made rigorous by being reduced to the algebra of inequalities, one must have both the algebra of inequalities, and facts about the concepts of the calculus that can be expressed in terms of the algebra of inequalities.

In the early nineteenth century, three conditions held for the first time: Rigor was considered important; there was a well-developed algebra of inequalities; and, certain properties were known about the basic concepts of analysis—limits, convergence, continuity, derivatives, integrals—properties which could be expressed in the language of inequalities if desired. Cauchy, followed by Riemann and Weierstrass, gave the calculus a rigorous basis, using the already-existing algebra of inequalities, and built a logically-connected structure of theorems about the concepts of the calculus. It is our task to explain how these three conditions—the developed algebra of inequalities, the importance of rigor, the appropriate properties of the concepts of the calculus—came to be.

The Algebra of Inequalities. Today, the algebra of inequalities is studied in calculus courses because of its use as a basis for the calculus, but why should it have been studied in the eighteenth century when this application was unknown? In the eighteenth century, inequalities were important in the study of a major class of results: approximations. For example, consider an equation such as $(x + 1)^{\mu} = a$, for μ not an integer. Usually a cannot be found exactly, but it can be approximated by an infinite series. In general, given some number n of terms of such an approximating series, eighteenth-century mathematicians sought to compute an upper bound on the error in the approximation—that is, the difference between the sum of the series and the nth partial sum. This computation was a problem in the algebra of inequalities. Jean d'Alembert solved it for the important case of the binomial series; given the number of terms of the series n, and assuming implicitly that the series converges to its sum, he could find the bounds on the error—that is, on the remainder of the series after the nth term—by bounding the series above and below with convergent geometric progressions [**7**]. Similarly, Joseph-Louis Lagrange invented a new approximation method using continued fractions and, by extremely intricate inequality-calculations, gave necessary and sufficient conditions for a given iteration of the approximation to be closer to the result than the previous iteration [**8**]. Lagrange also derived the Lagrange remainder of the Taylor series [**9**], using an inequality which bounded the remainder above and below by the maximum and minimum values of the nth derivative and then applying the intermediate-value theorem for continuous functions. Thus through such eighteenth-century work [**10**], there was by the end of the eighteenth century a developed algebra of inequalities, and people used to working with it. Given an n, these people are used to finding an error—that is, an epsilon.

Changing Attitudes toward Rigor. Mathematicians were much more interested in finding rigorous foundations for the calculus in 1800 than they had been a hundred years before. There are many reasons for this: no one enough by itself, but apparently sufficient when acting together. Of course one might think that eighteenth-century mathematicians were always making errors because of the lack of an explicitly-formulated rigorous foundation. But this did not occur. They were usually right, and for two reasons. One is that if one deals with real variables, functions of one variable, series which are power series, and functions arising from physical problems, errors will not occur too often. A second reason is that mathematicians like Euler and Laplace had a deep insight into the basic properties of the concepts of the calculus, and were able to choose fruitful methods and evade pitfalls. The only "error" they committed was to use methods that shocked mathematicians of later ages who had grown up with the rigor of the nineteenth century.

What then were the reasons for the deepened interest in rigor? One set of reasons was philosophical. In 1734, the British philosopher Bishop Berkeley had attacked the calculus on the ground that it was not rigorous. In *The Analyst, or a Discourse Addressed to an Infidel Mathematician*, he said that mathematicians had no business attacking the unreasonableness of

religion, given the way they themselves reasoned. He ridiculed fluxions—"velocities of evanescent increments"—calling the evanescent increments "ghosts of departed quantities" [**11**]. Even more to the point, he correctly criticized a number of specific arguments from the writings of his mathematical contemporaries. For instance, he attacked the process of finding the fluxion (our derivative) by reviewing the steps of the process: if we consider $y = x^2$, taking the ratio of the differences $((x + h)^2 - x^2)/h$, then simplifying to $2x + h$, then letting h vanish, we obtain $2x$. But is h zero? If it is, we cannot meaningfully divide by it; if it is not zero, we have no right to throw it away. As Berkeley put it, the quantity we have called h "might have signified either an increment or nothing. But then, which of these soever you make it signify, you must argue consistently with such its signification" [**12**].

Since an adequate response to Berkeley's objections would have involved recognizing that an equation involving limits is a shorthand expression for a sequence of inequalities—a subtle and difficult idea—no eighteenth-century analyst gave a fully adequate answer to Berkeley. However, many tried. Maclaurin, d'Alembert, Lagrange, Lazare Carnot, and possibly Euler, all knew about Berkeley's work, and all wrote something about foundations. So Berkeley did call attention to the question. However, except for Maclaurin, no leading mathematician spent much time on the question because of Berkeley's work, and even Maclaurin's influence lay in other fields.

Another factor contributing to the new interest in rigor was that there was a limit to the number of results that could be reached by eighteenth-century methods. Near the end of the century, some leading mathematicians had begun to feel that this limit was at hand. D'Alembert and Lagrange indicate this in their correspondence, with Lagrange calling higher mathematics "decadent" [**13**]. The philosopher Diderot went so far as to claim that the mathematicians of the eighteenth century had "erected the pillars of Hercules" beyond which it was impossible to go [**14**]. Thus, there was a perceived need to consolidate the gains of the past century.

Another "factor" was Lagrange, who became increasingly interested in foundations, and, through his activities, interested other mathematicians. In the eighteenth century, scientific academies offered prizes for solving major outstanding problems. In 1784, Lagrange and his colleagues posed the problem of foundations of the calculus as the Berlin Academy's prize problem. Nobody solved it to Lagrange's satisfaction, but two of the entries in the competition were later expanded into full-length books, the first on the Continent, on foundations: Simon L'Huilier's *Exposition élémentaire des principes des calculs supérieurs*, Berlin, 1787, and Lazare Carnot's *Réflexions sur la métaphysique du calcul infinitésimal*, Paris, 1797. Thus Lagrange clearly helped revive interest in the problem.

Lagrange's interest stemmed in part from his respect for the power and generality of algebra; he wanted to gain for the calculus the certainty he believed algebra to possess. But there was another factor increasing interest in foundations, not only for Lagrange, but for many other mathematicians by the end of the eighteenth century: the need to teach. Teaching forces one's attention to basic questions. Yet before the mideighteenth century, mathematicians had often made their living by being attached to royal courts. But royal courts declined; the number of mathematicians increased; and mathematics began to look useful. First in military schools and later on at the Ecole Polytechnique in Paris, another line of work became available: teaching mathematics to students of science and engineering. The Ecole Polytechnique was founded by the French revolutionary government to train scientists, who, the government believed, might prove useful to a modern state. And it was as a lecturer in analysis at the Ecole Polytechnique that Lagrange wrote his two major works on the calculus which treated foundations; similarly, it was 40 years earlier, teaching the calculus at the Military Academy at Turin, that Lagrange had first set out to work on the problem of foundations. Because teaching forces one to ask basic questions about the nature of the most important concepts, the change in the economic circumstances of mathematicians—the need to teach—provided a catalyst for the crystallization of the foundations of the calculus out of the historical and mathematical background. In fact, even well into the nineteenth century, much of foundations was born in the teaching situation; Weierstrass's

foundations come from his lectures at Berlin; Dedekind first thought of the problem of continuity while teaching at Zurich; Dini and Landau turned to foundations while teaching analysis; and, most important for our present purposes, so did Cauchy. Cauchy's foundations of analysis appear in the books based on his lectures at the Ecole Polytechnique; his book of 1821 was the first example of the great French tradition of *Cours d'analyse*.

The Concepts of the Calculus. Arising from algebra, the algebra of inequalities was now there for the calculus to be reduced to; the desire to make the calculus rigorous had arisen through consolidation, through philosophy, through teaching, through Lagrange. Now let us turn to the mathematical substance of eighteenth-century analysis, to see what was known about the concepts of the calculus before Cauchy, and what he had to work out for himself, in order to define, and prove theorems about, limit, convergence, continuity, derivatives, and integrals.

First, consider the concept of limit. As we have already pointed out, since Newton the limit had been thought of as a *bound* which could be approached closer and closer, though not surpassed. By 1800, with the work of L'Huilier and Lacroix on alternating series, the restriction that the limit be one-sided had been abandoned. Cauchy systematically translated this refined limit-concept into the algebra of inequalities, and used it in proofs once it had been so translated; thus he gave reality to the oft-repeated eighteenth-century statement that the calculus could be based on limits.

For example, consider the concept of convergence. Maclaurin had said already that the sum of a series was the limit of the partial sums. For Cauchy, this meant something precise. It meant that, given an ε, one could find n such that, for more than n terms, the sum of the infinite series is within ε of the nth partial sum. That is the reverse of the error-estimating procedure that d'Alembert had used. From his definition of a series having a sum, Cauchy could prove that a geometric progression with radius less in absolute value than 1 converged to its usual sum. As we have said, d'Alembert had shown that the binomial series for, say, $(1 + x)^{p/q}$ could be bounded above and below by convergent geometric progressions. Cauchy assumed that if a series of positive terms is bounded above, term-by-term, by a convergent geometric progression, then it converges; he then used such comparisons to prove a number of tests for convergence: the root test, the ratio test, the logarithm test. The treatment is quite elegant [**15**]. Taking a technique used a few times by men like d'Alembert and Lagrange on an ad hoc basis in approximations, and using the definition of the sum of a series based on the limit-concept, Cauchy created the first rigorous theory of convergence.

Let us now turn to the concept of continuity. Cauchy gave essentially the modern definition of continuous function, saying that the function $f(x)$ is continuous on a given interval if for each x in that interval "the numerical [i.e., absolute] value of the difference $f(x + \alpha) - f(x)$ decreases indefinitely with α" [**16**]. He used this definition in proving the intermediate-value theorem for continuous functions [**17**]. The proof proceeds by examining a function $f(x)$ on an interval, say $[b, c]$, where $f(b)$ is negative, $f(c)$ positive, and dividing the interval $[b, c]$ into m parts of width $h = (c - b)/m$. Cauchy considered the sign of the function at the points $f(b), f(b + h), \ldots,$ $f(b + (m - 1)h), f(c)$; unless one of the values of f is zero, there are two values of x differing by h such that f is negative at one, positive at the other. Repeating this process for new intervals of width $(c - b)/m, (c - b)/m^2, \ldots,$ gives an increasing sequence of values of x: $b, b_1, b_2, \ldots$ for which f is negative, and a decreasing sequence of values of x: $c, c_1, c_2, \ldots$ for which f is positive, and such that the difference between b_k and c_k goes to zero. Cauchy asserted that these two sequences must have a common limit a. He then argued that since $f(x)$ is continuous, the sequence of negative values $f(b_k)$ and of positive values $f(c_k)$ both converge toward the common limit $f(a)$, which must therefore be zero.

Cauchy's proof involves an already existing technique, which Lagrange had applied in approximating real roots of polynomial equations. If a polynomial was negative for one value of the variable, positive for another, there was a root in between, and the difference between those two values of the variable bounded the error made in taking either as an approximation to the

root [18]. Thus again we have the algebra of inequalities providing a technique which Cauchy transformed from a tool of approximation to a tool of rigor.

It is worth remarking at this point that Cauchy, in his treatment both of convergence and of continuity, implicitly assumed various forms of the completeness property for the real numbers. For instance, he treated as obvious that a series of positive terms, bounded above by a convergent geometric progression, converges: also, his proof of the intermediate-value theorem assumes that a bounded monotone sequence has a limit. While Cauchy was the first systematically to exploit inequality proof techniques to prove theorems in analysis, he did not identify all the implicit assumptions about the real numbers that such inequality techniques involve. Similarly, as the reader may have already noticed, Cauchy's definition of continuous function does not distinguish between what we now call point-wise and uniform continuity; also, in treating series of functions, Cauchy did not distinguish between pointwise and uniform convergence. The verbal formulations like "for all" that are involved in choosing deltas did not distinguish between "for any epsilon and for all x" and "for any x, given any epsilon" [19]. Nor was it at all clear in the 1820's how much depended on this distinction, since proofs about continuity and convergence were in themselves so novel. We shall see the same confusion between uniform and point-wise convergence as we turn now to Cauchy's theory of the derivative.

Again we begin with an approximation. Lagrange gave the following inequality about the derivative:

$$f(x+h)=f(x)+hf'(x)+hV, \tag{2}$$

where V goes to 0 with h. He interpreted this to mean that, given any D, one can find h sufficiently small so that V is between $-D$ and $+D$ [20]. Clearly this is equivalent to (1) above, Cauchy's delta-epsilon characterization of the derivative. But how did Lagrange obtain this result? The answer is surprising; for Lagrange, formula (2) was a consequence of Taylor's theorem. Lagrange believed that any function (that is, any analytic expression, whether finite or infinite, involving the variable) had a unique power-series expansion (except possibly at a finite number of isolated points). This is because he believed that there was an "algebra of infinite series," an algebra exemplified by work of Euler such as the example we gave above. And Lagrange said that the way to make the calculus rigorous was to reduce it to algebra. Although there is no "algebra" of infinite series that gives power-series expansions without any consideration of convergence and limits, this assumption led Lagrange to define $f'(x)$ without reference to limits, as the coefficient of the linear term in h in the Taylor series expansion for $f(x+h)$. Following Euler, Lagrange then said that, for any power series in h, one could take h sufficiently small so that any given term of the series exceeded the sum of all the rest of the terms following it; this approximation, said Lagrange, is assumed in applications of the calculus to geometry and mechanics [21]. Applying this approximation to the linear term in the Taylor series produces (2), which I call the Lagrange property of the derivative. (Like Cauchy's (1), the inequality-translation Lagrange gives for (2) assumes that, given any D, one finds h sufficiently small so $|V| \leq D$ with no mention whatever of x.)

Not only did Lagrange state property (2) and the associated inequalities, he used them as a basis for a number of proofs about derivatives: for instance, to prove that a function with positive derivative on an interval is increasing there, to prove the mean-value theorem for derivatives, and to obtain the Lagrange remainder for the Taylor series. (Details may be found in the works cited in [22].) Lagrange also applied his results to characterize the properties of maxima and minima, and orders of contact between curves.

With a few modifications, Lagrange's proofs are valid—provided that property (2) can be justified. Cauchy borrowed and simplified what are in effect Lagrange's inequality proofs about derivatives, with a few improvements, basing them on his own (1). But Cauchy made these proofs legitimate because Cauchy defined the derivative precisely to satisfy the relevant inequalities. Once again, the key properties come from an approximation. For Lagrange, the derivative was

exactly—no epsilons needed—the coefficient of the linear term in the Taylor series; formula (2), and the corresponding inequality that $f(x+h)-f(x)$ lies between $h(f'(x) \pm D)$, were approximations. Cauchy brought Lagrange's inequality properties and proofs together with a definition of derivative devised to make those techniques rigorously founded [**22**].

The last of the concepts we shall consider, the integral, followed an analogous development. In the eighteenth century, the integral was usually thought of as the inverse of the differential. But sometimes the inverse could not be computed exactly, so men like Euler remarked that the integral could be approximated as closely as one liked by a sum. Of course, the geometric picture of an area being approximated by rectangles, or the Leibnizian definition of the integral as a sum, suggests this immediately. But what is important for our purposes is that much work was done on approximating the values of definite integrals in the eighteenth century, including considerations of how small the subintervals used in the sums should be when the function oscillates to a greater or lesser extent. For instance, Euler treated sums of the form $\Sigma_{k=0}^{n} f(x_k)(x_{k+1}-x_k)$ as approximations to the integral $\int_{x_0}^{x_n} f(x)\,dx$ [**23**].

In 1820, S.-D. Poisson, who was interested in complex integration and therefore more concerned than most people about the existence and behavior of integrals, asked the following question. If the integral F is defined as the antiderivative of f, and if $b-a=nh$, can it be proved that $F(b)-F(a)=\int_a^b f(x)\,dx$ is the limit of the sum

$$S = hf(a) + hf(a+h) + \cdots + hf(a+(n-1)h)$$

as h gets small? (S is an approximating sum of the eighteenth-century sort.) Poisson called this result "the fundamental proposition of the theory of definite integrals." He proved it by using another inequality-result: the Taylor series with remainder. First, he wrote $F(b)-F(a)$ as the telescoping sum

$$(3) \qquad F(a+h)-F(a)+F(a+2h)-F(a+h)+\cdots+F(b)-F(a+(n-1)h).$$

Then, for each of the terms of the form $F(a+kh)-F(a+(k-1)h)$, Taylor's series with remainder gives, since by definition $F' = f$,

$$F(a+kh)-F(a+(k-1)h) = hf(a+(k-1)h) + R_k h^{1+w}$$

where $w>0$, for some R_k. Thus the telescoping sum (3) becomes

$$hf(a) + hf(a+h) + \cdots + hf(a+(n-1)h) + (R_1 + \cdots + R_n)h^{1+w}.$$

So $F(b)-F(a)$ and the sum S differ by $(R_1+\cdots+R_n)h^{1+w}$. Letting R be the maximum value for the R_k,

$$(R_1+\cdots+R_n)h^{1+w} \le n \cdot R(h^{1+w}) = R \cdot nh \cdot h^w = R(b-a)h^w.$$

Therefore, if h is taken sufficiently small, $F(b)-F(a)$ differs from S by less than any given quantity [**24**].

Poisson's was the first attempt to prove the equivalence of the antiderivative and limit-of-sums conceptions of the integral. However, besides the implicit assumptions of the existence of antiderivatives and bounded first derivatives for f on the given interval, the proof assumes that the subintervals on which the sum is taken are all equal. Should the result not hold for unequal divisions also? Poisson thought so, and justified it by saying, "If the integral is represented by the area of a curve, this area will be the same, if we divide the difference... into an infinite number of equal parts, or an infinite number of unequal parts following any law" [**25**]. This, however, is an assertion, not a proof. And Cauchy saw that a proof was needed.

Cauchy did not like formalistic arguments in supposedly rigorous subjects, saying that most algebraic formulas hold "only under certain conditions, and for certain values of the quantities they contain" [**26**]. In particular, one could not assume that what worked for finite expressions automatically worked for infinite ones. Thus, Cauchy showed that the sum of the series

$1/1 + 1/4 + 1/9 + \cdots$ was $\pi^2/6$ by actually calculating the difference between the nth partial sum and $\pi^2/6$ and showing that it was arbitrarily small [**27**]. Similarly, just because there was an operation called taking a derivative did not mean that the inverse of that operation always produced a result. The existence of the definite integral had to be proved. And how does one prove existence in the 1820's? One constructs the mathematical object in question by using an eighteenth-century approximation that converges to it. Cauchy defined the integral as the limit of Euler-style sums $\Sigma f(x_k)(x_{k+1} - x_k)$ for $x_{k+1} - x_k$ sufficiently small. Assuming explicitly that $f(x)$ was continuous on the given interval (and implicitly that it was uniformly continuous), Cauchy was able to show that all sums of that form approach a fixed value, called by definition the integral of the function on that interval. This is an extremely hard proof [**28**]. Finally, borrowing from Lagrange the mean-value theorem for integrals, Cauchy proved the Fundamental Theorem of Calculus [**29**].

Conclusion. Here are all the pieces of the puzzle we originally set out to solve. Algebraic approximations produced the algebra of inequalities; eighteenth-century approximations in the calculus produced the useful properties of the concepts of analysis: d'Alembert's error-bounds for series, Lagrange's inequalities about derivatives, Euler's approximations to integrals. There was a new interest in foundations. All that was needed was a sufficiently great genius to build the new foundation.

Two men came close. In 1816, Carl Friedrich Gauss gave a rigorous treatment of the convergence of the hypergeometric series, using the technique of comparing a series with convergent geometric progressions; however, Gauss did not give a general foundation for all of analysis. Bernhard Bolzano, whose work was little known until the 1860's, echoing Lagrange's call to reduce the calculus to algebra, gave in 1817 a definition of continuous function like Cauchy's and then proved—by a different technique from Cauchy's—the intermediate-value theorem [**30**]. But it was Cauchy who gave rigorous definitions and proofs for all the basic concepts; it was he who realized the far-reaching power of the inequality-based limit concept; and it was he who gave us—except for a few implicit assumptions about uniformity and about completeness—the modern rigorous approach to calculus.

Mathematicians are used to taking the rigorous foundations of the calculus as a completed whole. What I have tried to do as a historian is to reveal what went into making up that great achievement. This needs to be done, because completed wholes by their nature do not reveal the separate strands that go into weaving them—especially when the strands have been considerably transformed. In Cauchy's work, though, one trace indeed was left of the origin of rigorous calculus in approximations—the letter epsilon. The ε corresponds to the initial letter in the word "erreur" (or "error"), and Cauchy in fact used ε for "error" in some of his work on probability [**31**]. It is both amusing and historically appropriate that the "ε," once used to designate the "error" in approximations, has become transformed into the characteristic symbol of precision and rigor in the calculus. As Cauchy transformed the algebra of inequalities from a tool of approximation to a tool of rigor, so he transformed the calculus from a powerful method of generating results to the rigorous subject we know today.

References

1. A. -L. Cauchy, Cours d'analyse, Paris, 1821; in Oeuvres complètes d'Augustin Cauchy, series 2, vol. 3, Paris, Gauthier-Villars, 1899, p. 19.

2. A. -L. Cauchy, Résumé des leçons données à l'école royale polytechnique sur le calcul infinitésimal, Paris, 1823; in Oeuvres, series 2, vol. 4, p. 44. Cauchy used i for the increment; otherwise the notation is his.

3. Isaac Newton, Mathematical Principles of Natural Philosophy, 3rd ed., 1726, tr. A. Motte, revised by Florian Cajori, University of California Press, Berkeley, 1934, Scholium to Lemma XI, p. 39.

4. Johann Bernoulli, Opera Omnia, IV, 8; section entitled "De seriebus varia, Corollarium III," cited by D. J. Struik, A Source Book in Mathematics, 1200–1800, Harvard, Cambridge, 1969, p. 321.

5. Boyer, History of Mathematics, p. 487; Euler's paper is in Comm. Acad. Sci. Petrop., 7, 1734–5, pp. 123–34;

in Leonhard Euler, Opera omnia, series 1, vol. 14, pp. 73–86.

6. J. V. Grabiner, The Origins of Cauchy's Rigorous Calculus, M. I. T. Press, Cambridge and London, 1981, chapter 2.

7. J. d'Alembert, Réflexions sur les suites et sur les racines imaginaires, in Opuscules mathématiques, vol. 5, Briasson, Paris, 1768, pp. 171–215; see especially pp. 175–178.

8. J. -L. Lagrange, Traité de la résolution des équations numériques de tous les degrés, 2nd ed., Courcier, Paris, 1808; in Oeuvres de Lagrange, Gauthier-Villars, Paris, 1867–1892, vol. 8, pp. 162–163.

9. Lagrange, Théorie des fonctions analytiques, 2nd ed., Paris, 1813, in Oeuvres, vol. 9, pp. 80–85; compare Lagrange, Leçons sur le calcul des fonctions, Paris, 1806, in Oeuvres, vol. 10, pp. 91–95.

10. Grabiner, Origins of Cauchy's Rigorous Calculus, pp. 56–68; compare H. Goldstine, A History of Numerical Analysis from the 16th through the 19th Century, Springer-Verlag, New York, Heidelberg, Berlin, 1977, chapters 2–4.

11. George Berkeley, The Analyst, section 35.

12. Analyst, section 15. Berkeley used the function x^n where we have used x^2, and a Newtonian notation, lower-case o, for the increment.

13. Letter from Lagrange to d'Alembert, 24 February 1772, in Oeuvres de Lagrange, vol. 13, p. 229.

14. D. Diderot, De l'interprétation de la nature, in Oeuvres philosophiques, ed., P. Vernière, Garnier, Paris, 1961, pp. 180–181.

15. Cauchy, Cours d'analyse, Oeuvres, series 2, vol. 3; for real-valued series, see especially pp. 114–138.

16. Cauchy, op. cit., p. 43. So did Bolzano; see below, and note 30.

17. Cauchy, op. cit., pp. 378–380. For an English translation of this proof, see Grabiner, Origins, pp. 167–168. For clarity, I have substituted $b, b_1, b_2, \ldots$ and $c, c_1, c_2, \ldots$ for Cauchy's $x_0, x_1, x_2, \ldots$ and $X, X', X'', \ldots$ in the present version.

18. Lagrange, Equations numériques, sections 2 and 6, in Oeuvres, vol. 8; also in Lagrange, Leçons élémentaires sur les mathématiques données à l'école normale en 1795, Séances des Ecoles Normales, Paris, 1794–1795; in Oeuvres, vol. 7, pp. 181–288; this method is on pp. 260–261.

19. I. Grattan-Guinness, Development of the Foundations of Mathematical Analysis from Euler to Riemann, M. I. T. Press, Cambridge and London, 1970, p. 123, puts it well: "Uniform convergence was tucked away in the word "always," with no reference to the variable at all."

20. Lagrange, Leçons sur le calcul des fonctions, Oeuvres 10, p. 87; compare Lagrange, Théorie des fonctions analytiques, Oeuvres 9, p. 77. I have substituted h for the i Lagrange used for the increment.

21. Lagrange, Théorie des fonctions analytiques, Oeuvres 9, p. 29. Compare Leçons sur le calcul des fonctions, Oeuvres 10, p. 101. For Euler, see his Institutiones calculi differentialis, St. Petersburg, 1755; in Opera, series 1, vol. 10, section 122.

22. Grabiner, Origins of Cauchy's Rigorous Calculus, chapter 5; also J. V. Grabiner, The origins of Cauchy's theory of the derivative, Hist. Math., 5, 1978, pp. 379–409.

23. The notation is modernized. For Euler, see Institutiones calculi integralis, St. Petersburg, 1768–1770, 3 vols; in Opera, series 1, vol. 11, p. 184. Eighteenth-century summations approximating integrals are treated in A. P. Iushkevich, O vozniknoveniya poiyatiya ob opredelennom integrale Koshi, Trudy Instituta Istorii Estestvoznaniya, Akademia Nauk SSSR, vol. 1, 1947, pp. 373–411.

24. S. D. Poisson, Suite du mémoire sur les intégrales définies, Journ. de l'Ecole polytechnique, Cah. 18, 11, 1820, pp. 295–341, 319–323. I have substituted h, w for Poisson's α, k, and have used R_1 for his R_0.

25. Poisson, op. cit., pp. 329–330.

26. Cauchy, Cours d'analyse, Introduction, Oeuvres, Series 2, vol. 3, p. iii.

27. Cauchy, Cours d'analyse, Note VIII, Oeuvres, series 2, vol. 3, pp. 456–457.

28. Cauchy, Calcul infinitésimal, Oeuvres, series 2, vol. 4, 122–25; in Grabiner, Origins of Cauchy's Rigorous Calculus, pp. 171–175 in English translation.

29. Cauchy, op. cit., pp. 151–152.

30. B. Bolzano, Rein analytischer Beweis des Lehrsatzes dass zwischen je zwey Werthen, die ein entgegengesetztes Resultat gewaehren, wenigstens eine reele Wurzel der Gleichung liege, Prague, 1817. English version, S. B. Russ, A translation of Bolzano's paper on the intermediate value theorem, Hist. Math., 7, 1980, pp. 156–185. The contention by Grattan-Guinness, Foundations, p. 54, that Cauchy took his program of rigorizing analysis, definition of continuity, Cauchy criterion, and proof of the intermediate-value theorem, from Bolzano's paper without acknowledgement is not, in my opinion, valid; the similarities are better explained by common prior influences, especially that of Lagrange. For a documented argument to this effect, see J. V. Grabiner, Cauchy and Bolzano: Tradition and transformation in the history of mathematics, to appear in E. Mendelsohn, Transformation and Tradition in the Sciences, Cambridge University Press, Cambridge, forthcoming; see also Grabiner, Origins of Cauchy's Rigorous Calculus, pp. 69–75, 102–105, 94–96, 52–53.

31. Cauchy, Sur la plus grande erreur à craindre dans un résultat moyen, et sur le système de facteurs qui rend cette plus grande erreur un minimum, Comptes rendus 37, 1853; in Oeuvres, series 1, vol. 12, pp. 114–124.

CENTER SECTION

(Vol. 90, No. 3, March 1983)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P, L. Faces of Science. V.V. Nalimov. ISI Pr, 1981, xv + 297 pp, $22.50. [ISBN: 0-89495-010-X] A prominent Soviet statistician contemplates the social nature of science, viewing it as a large self-organized information system, a cybernetic system. This strongly humanistic analysis focusses especially on the role of language and of uncertainty in shaping scientific theory and experimental practice. GHM

General. Practical Math Handbook for the Building Trades. Paul Calter. Prentice-Hall, 1983, x + 294 pp, $8.95 (P). [ISBN: 0-13-692228-7] Very much a practical handbook at the precalculus level. Contains a general mathematical formula and table section (166 pages), including an outline of pocket calculator use, and a nice (110 page) section on construction including very complete and accessible tables and formulas for plumbing, heating, and financial computations. Handy for householders and also as a source of "practical" problems for high school algebra. JAS

General, S, L*. Tools for Teaching. Birkhauser Boston, 1981. UMAP Modules 1977-1979. xii + 727 pp, $35 [ISBN: 3-7643-3049-X]; UMAP Modules for 1980. xii + 690 pp, $35. [ISBN: 3-7643-3059-7] Annual collections of field-tested modules from project UMAP. These lecture-sized supplements offer diverse opportunities for classroom enrichment: cardiac output, Monte Carlo experiments, digestion in sheep, scheduling prison guards, et al. LAS

General, S, L*. Handbook of Cubik Math. Alexander H. Frey, Jr., David Singmaster. Enslow Pub, 1982, viii + 193 pp, $9.95 (P). [ISBN: 0-89490-058-7] A well-written elaboration of Singmaster's Notes on Rubik's Magic Cube (TR, February 1982), containing much the same material in more accessible form: notation, solution methods, analysis of basic moves, subgroup structures, advanced methods. Deals only with the 3x3 cube. LAS

General, S, L*. Inside Rubik's Cube and Beyond. Christoph Bandelow. Birkhauser Boston, 1982, 125 pp, $3.95 (P). [ISBN: 3-7643-3078-3] A systematic, mathematical treatment of Rubik's cube and its associated group theory, complete with definitions, theorems and proofs. The notation is similar but not identical to the now-standard notation introduced by Singmaster. Includes discussion of other similar puzzles (tetrahedron, dodecahedron, 2x2 cube, but not the 4x4 cube), flowchart for a computer program and a catalogue of maneuvers. LAS

General, P, L.** An Assessment of Research-Doctorate Programs in the United States: Mathematical & Physical Sciences. Ed: Lyle V. Jones, Gardner Lindzey, Porter E. Coggeshall. National Academy Pr, 1982, xii + 243 pp, $10.50 (P). [ISBN: 0-309-03299-7] Full report on the recent, widely-publicized survey of characteristics and quality in Ph.D. programs. Data from 16 different measures (of size, reputation, library, graduates, publications, research support) are presented in both raw, standardized, and graphical forms. LAS

Mathematics Appreciation, S*(13-15), L. Mathematical Fallacies and Paradoxes. Bryan H. Bunch. Van Nostrand Reinhold, 1982, xi + 216 pp, $16.95. [ISBN: 0-442-24905-5] A well-written elementary introduction to paradoxes of logic, geometry, and algebra, assuming only the rudiments of high school algebra and geometry. Moves carefully but swiftly from traditional fallacies (e.g., division by zero) through liar paradoxes to more sophisticated things such as Gödel's theorem, Banach-Tarski paradox and Zeno's paradoxes. LAS

Education, S*(13). Studying Mathematics. Mary Catharine Hudspeth, Lewis R. Hirsch. Kendall/Hunt, 1982, xi + 51 pp, $4.95 (P). [ISBN: 0-8403-2768-4] Nearly everything you want students to know about preparing for and being successful in elementary courses. Covers topics such as: how to read a math text, preparing for exams, organized note-taking. Responds well to standard student complaints/excuses, e.g., "I would have passed if I didn't make so many dumb mistakes." JRG

Education, P. Pushbutton Mathematics: Calculator Math Problems, Examples, and Activities. Kenneth P. Goldberg. Prentice-Hall, 1982, x + 197 pp, $8.95 (P). [ISBN: 0-13-743302-6] A guide to the use of calculators in the mathematics classroom. Gives detailed examples of some topics and then additional activities in seven areas: elementary algebra, trigonometry, intermediate/advanced algebra, geometry, business/consumer mathematics, probability and statistics, and elementary calculus. RSK

Education, P*, L. The Future of College Mathematics. Ed: Anthony Ralston, Gail S. Young. Springer-Verlag, 1983, ix + 278 pp. [ISBN: 0-387-90813-7] Proceedings of a 1982 summer conference at Williams College devoted to an analysis of the proposed need for a freshman/sophomore discrete mathematics alternative to calculus. Includes both papers--suggested course outlines, interface with client disciplines, problems of implementation--and summaries of discussion. LAS

History. Gabor Szegő: Collected Papers. Ed: Richard Askey. Birkhauser Boston, 1982, $180 set. Volume 1: 1915-1927, xx + 857 pp [ISBN: 3-7643-3056-2]; Volume 2: 1927-1943, x + 869 pp [ISBN: 3-7643-3060-0]; Volume 3: 1945-1972, x + 880 pp. [ISBN: 3-7643-3061-9] All of Szegő's papers, arranged in chronological order, introduced by a short biography; commentary on Szegő's life and work by G. Polya, P.C. Rosenbloom, and R. Askey; reviews of three of Szegő's books; and two mathematical commentaries on Szegő's work. LAS

History, S*(14-18), P*, L*.** Neyman--from life. Constance Reid. Springer-Verlag, 1982, 298 pp, $19.80. [ISBN: 0-387-90747-5] Fascinating non-technical account of the life of Jerzy Neyman, one of the great pioneers of mathematical statistics, by the author of Hilbert and Courant in Göttingen and New York. Written while Neyman (1894-1981) was still alive, it contains his personal recollections as well as those of his colleagues, including his best-known collaborator, Egon Pearson. Must reading for anyone interested in the history of statistics. RSK

History, L*. The Collected Letters of Colin MacLaurin. Ed: Stella Mills. Shiva Pub, 1982, xx + 496 pp, $35. [ISBN: 0-906812-08-9] Colin MacLaurin, one of Scotland's greatest mathematicians, helped interpret and defend Newton in, e.g., his 1742 Treatise of Fluxions, an answer to Bishop Berkeley's 1734 The Analyst. This volume of letters, both general and scientific, traces the development of his ideas in exchanges with Newton and Stirling and many less well known eighteenth century scientists and mathematicians. LAS

Foundations, P. Mathematical Logic, Revised Edition. Willard Van Orman Quine. Harvard U Pr, 1976, xii + 346 pp, $25; $9.95 (P). [ISBN: 0-674-55450-7] Unaltered reprinting of 1951 revised edition. A classic of the field, but now considerably dated as to notation and emphasis. GHM

Foundations, T*(15-16: 1), S, L*. Formal Number Theory and Computability: A Workbook. Alec Fisher. Clarendon Pr, 1982, xiii + 190 pp, $12.50 (P); $37. [ISBN: 0-19-853188-5; 0-19-853178-8] Interesting text for first undergraduate course in mathematical logic. Focuses on Gödel's incompleteness theorem and related consistency and decidability results. Covers general propositional and predicate calculus briefly. Introduces computability using register machines. Numerous exercises; complete solutions in appendix. KS

Graph Theory, P. Zero-Symmetric Graphs: Trivalent Graphical Regular Representations of Groups. H.S.M. Coxeter, Roberto Frucht, David L. Powers. Academic Pr, 1981, ix + 170 pp, $15. [ISBN: 0-12-194580-4] Description of all known zero-symmetric graphs with not more than 120 vertices. Zero-symmetric graphs are finite trivalent graphs whose automorphism group acts regularly on the vertices. KS

Combinatorics, P. Lecture Notes in Mathematics-952: Combinatorial Mathematics IX. Ed: Elizabeth J. Billington, Sheila Oates-Williams, Anne Penfold Street. Springer-Verlag, 1982, xi + 443 pp, $23 (P). [ISBN: 0-387-11601-X] Proceedings of the Ninth Australian Conference on Combinatorial Mathematics held at the University of Brisbane, Australia, August 1981. Texts of seven of the nine invited talks, plus 20 contributed papers. LAS

Discrete Mathematics, T(17-18: 1, 2), S, L. Introduction to Coding Theory. J.H. van Lint. Grad. Texts in Math., No. 86. Springer-Verlag, 1982, ix + 171 pp, $24. [ISBN: 0-387-11284-7] Graduate level text emphasizing construction and analysis of codes with some attention to mathematical problems of decoding. Assumes knowledge of abstract and linear algebra, combinatorics, and probability theory. KS

Number Theory, S*(13), P, L*. Solving Equations in Integers. A.O. Gelfond. Trans: O.B. Sheinin. Little Math. Lib. MIR Pub, 1981, 56 pp, $2 (P). Based on a lecture the author delivered in 1951 for participants in a Mathematical Olympiad. Focuses on classical results concerning equations in two or three unknowns of low degree. CEC

Number Theory, T(17: 1), S, P, L. Adeles and Algebraic Groups. A. Weil. Progress in Math., V. 23. Birkhauser Boston, 1982, 126 pp, $10. [ISBN: 3-7643-3092-9] This volume contains the original lecture notes (from 1959-60) in which the concept of adeles was first introduced. The notes have been supplemented by an extended bibliography and a brief survey of subsequent research. CEC

Number Theory, S(18), P. Periods of Hilbert Modular Surfaces. Takayuki Oda. Progress in Math., V. 19. Birkhauser Boston, 1982, xvi + 123 pp, $10. [ISBN: 3-7643-3084-8] The author investigates Hodge structures of Hilbert modular surfaces and shows that the Hodge structure attached to a Hilbert modular surface is a direct sum of tensor products of two Hodge structures of weight 1. Includes a

bibliography. CEC

Number Theory, P. Ternary Quadratic Forms and Norms. Ed: Olga Taussky. Lect. Notes in Pure & Appl. Math., V. 79. Dekker, 1982, vii + 135 pp, $19.75 (P). [ISBN: 0-8247-1651-5] Presents six papers given at a Cal Tech seminar during the 1974-75 academic year. Of special interest is Zassenhaus's exposition of Gauss's theory of ternary forms. Other contributors are Kisilevsky, Plesken, Rehm, Taussky, Weiss, and via letter, van der Waerden. SG

Number Theory, T(15: 1), S, P, L. The Higher Arithmetic: An Introduction to the Theory of Numbers, Fifth Edition. H. Davenport. Cambridge U Pr, 1982, 189 pp, $22.95; $9.95 (P). [ISBN: 0-521-24422-6; 0-521-28678-6] A new edition of this classic work, which was first published in 1952 and last reissued in 1970. This edition includes problems and solutions and occasional updates. CEC

Linear Algebra, P. Recent Applications of Generalized Inverses. Ed: S.L. Campbell. Research Notes in Math., No. 66. Pitman Pub, 1982, 274 pp, $23.95 (P). [ISBN: 0-273-08550-6] Collection of papers designed to reflect current trends in research done on and with generalized inverses of matrices. Includes applications to electrical networks and Markov chains, use in numerical analysis, and extension of theory to matrices over rings and infinite dimensional spaces. KS

Linear Algebra, T(16-17: 1, 2), S, P, L. Cartesian Tensors: With Applications to Mechanics, Fluid Mechanics and Elasticity. A.M. Goodbody. Math. & Its Appl. Halsted Pr, 1982, 298 pp, $75. [ISBN: 0-470-27254-6] Multilinear algebra developed with physical applications--especially mechanics--in mind. Emphasis on computation; many explicit calculations, examples. A readable text for engineering students with some mathematical sophistication. Exercises, many with solutions. PZ

Linear Algebra, T(13: 1), S, L. Matrix Algebra Useful for Statistics. Shayle R. Searle. Wiley, 1982, xxii + 438 pp, $29.95. [ISBN: 0-471-86681-4] The final two chapters on regression and linear models in statistics are described using matrix terminology. The objective of the book is to lead even the minimally prepared student (high school algebra) to a familiarity with the notation and an understanding of the relevant theory (including generalized inverses, partitioned matrices, eigenvalues and eigenvectors). The theoretical development is profusely illustrated with numerical examples and applications from the applied sciences. LCL

Algebra, P. Lecture Notes in Mathematics-924: Séminaire d'Algèbre Paul Dubreil et Marie-Paule Malliavin. Ed: M.-P. Malliavin. Springer-Verlag, 1982, v + 461 pp, $21.60 (P). [ISBN: 0-387-11496-3] Proceedings of a 1981 conference which took place in Paris. CEC

Algebra, T(18), S, P. A Concrete Approach to Division Rings. John Dauns. Res. & Educ. in Math., No. 2. Heldermann Verlag, 1982, xx + 417 pp, 78 DM (P). [ISBN: 3-88538-202-4] This is the first book to treat all types of division rings. By constructing and drawing attention to concrete examples, the author is able to lead even the non-expert easily and quickly to the frontiers of the subject--in marked contrast to the classical references on division rings in which the relevant facts are often given as corollaries to long chains of previous results. LCL

Algebra, T(18), P. Cohomology of Groups. Kenneth S. Brown. Grad. Texts in Math., No. 87. Springer-Verlag, 1982, x + 306 pp, $28. [ISBN: 0-387-90688-6] An introduction to a subject which draws on a rich interplay between algebra and topology. The author builds the subject from the bottom, assuming only a knowledge of the first year graduate courses in algebra and algebraic topology. The final four chapters are more specialized. LCL

Algebra, P. Lecture Notes in Mathematics-944: Representations of Algebras. Ed: M. Auslander, E. Lluis. Springer-Verlag, 1982, v + 258 pp, $14 (P). [ISBN: 0-387-11577-3] Notes of four series of lectures given at the workshop held the week before the Third International Conference on Representations of Algebras held August 8-16, 1980 in Puebla, Mexico. The proceedings of this conference appeared as Lecture Notes 903. JAS

Algebra, P. Lecture Notes in Mathematics-951: Advances in Non-Commutative Ring Theory. Ed: P.J. Fleury. Springer-Verlag, 1982, 142 pp, $8.50 (P). [ISBN: 0-387-11597-8] Proceedings of the Twelfth George H. Hudson Symposium held at the State University College of Arts and Science at Plattsburgh, New York on April 23-25, 1981. JAS

Algebra, T(14-16). A First Undergraduate Course in Abstract Algebra, Third Edition. Abraham P. Hillman, Gerald L. Alexanderson. Wadsworth Pub, 1983, ix + 501 pp. [ISBN: 0-534-01195-0] The major changes from previous editions include several sections on discrete mathematics and a new chapter on coding theory. However, the book remains a solid introduction to modern algebra suitable for any student of mathematics, computer science, or statistics. (First Edition, Extended Review, August-September 1974; Second Edition, TR, June-July 1978.) SG

Algebra, P. Lecture Notes in Computer Science-144: Computer Algebra. Ed: Jacques Calmet. Springer-Verlag, 1982, xiv + 301 pp, $14 (P). [ISBN: 0-387-11607-9] Over thirty papers from the European Computer Algebra Conference held in Marseille, France in April 1982, dealing with the design, analysis, and implementation of constructive methods for solving various problems (e.g., root finding, factorization) in different algebraic structures. LCL

Algebra, P. Lecture Notes in Mathematics-933: Lie Algebras and Related Topics. Ed: D. Winter. Springer-Verlag, 1982, vi + 236 pp, $11 (P). [ISBN: 0-387-11563-3] Proceedings of a conference held

at Rutgers University at New Brunswick, New Jersey, May 29-31, 1981. JAS

Calculus, S(13). Areas and Logarithms. A.I. Markushevich. Trans: I. Aleksanova. MIR Pub, 1981, 70 pp, $2 (P). Integral as area. Develops logarithm as integral and derives properties, including series approximation for ln(1+x). Brief discussion of distribution of primes and relationship to logarithm function. Supplementary treatment of Simpson's Rule. JRG

Real Analysis, P. Lecture Notes in Mathematics-945: Measure Theory, Oberwolfach 1981. Ed: D. Kölzow, D. Maharam-Stone. Springer-Verlag, 1982, xv + 431 pp, $20.50 (P). [ISBN: 0-387-11580-3] Proceedings of a June 1981 conference at Oberwolfach. Concludes with a short list of open problems. LAS

Complex Analysis, S(18), P. Approximation Uniforme Qualitative sur des Ensembles non Bornés. Paul M. Gauthier, Walter Hengartner. Pr U Montreal, 1982, 89 pp, $11 (P). [ISBN: 2-7606-0574-4] Notes from lectures given at the August 1981 NATO seminar at Montreal. Surveys results and methods of locally uniform approximation theory for holomorphic functions on closed domains in Riemann surfaces, and for harmonic functions on domains in R^n. Approximation theorems of Runge and Mergelyan are generalized. PZ

Complex Analysis, P. Lecture Notes in Mathematics-950: Complex Analysis. Ed: J. Eells. Springer-Verlag, 1982, 428 pp, $20.50 (P). [ISBN: 0-387-11596-X] Proceedings of the 1980 international summer school in theoretical physics in Trieste: seven expositions on such topics as twistor theory, holomorphic vector bundles, and holomorphic foliations. LAS

Differential Equations, T(15: 1), S, L*.** A First Course in Differential Equations with Applications, Second Edition. Dennis G. Zill. Prindle, Weber & Schmidt, 1982, ix + 527 pp. [ISBN: 0-534-98011-2] Several revisions have been made in this edition, including over 600 new problems and many new examples and figures. All of the standard topics are covered, along with an introduction to partial differential equations, in this readable text. (TR, First Edition, August-September 1980.) CEC

Differential Equations, S(17-18), P. Methods of Bifurcation Theory. Shui-Nee Chow, Jack K. Hale. Grund. der math. Wissenschaften, B. 251. Springer-Verlag, 1982, xv + 515 pp, $48. [ISBN: 0-387-90664-9] The primary objective of this monograph is a presentation of those aspects of static and dynamic bifurcation theory which are of particular value in the study of differential equations. Introductory material on nonlinear functional analysis and the qualitative theory of differential equations makes the book accessible to nonspecialists. AO

Differential Equations, T(16-17: 1, 2), L. Ordinary Differential Equations, Second Edition. Philip Hartman. Birkhauser Boston, 1982, xv + 612 pp, $29.95. [ISBN: 3-7643-3068-6] Reprint of the 1973 corrected reprint (TR, June-July 1974) of the 1964 Wiley original edition. AO

Differential Equations, T(18: 1), S, P. Variational Principles and Free-Boundary Problems. Avner Friedman. Pure & Appl. Math. Wiley, 1982, ix + 710 pp, $52.50. [ISBN: 0-471-86849-3] A major effort to present in "a systematic and self-contained manner" many of the recent developments in the theory of free boundary problems. The first two chapters are used to assemble the necessary theory of partial differential operators with the remaining three aimed toward applications; jets and cavities, variational problems, and non-variational problems. Bibliography remarks, references, index. JS

Numerical Analysis, P, L. Applied Iterative Methods. Louis A. Hageman, David M. Young. Comp. Sci. & Appl. Math. Academic Pr, 1981, xvii + 386 pp, $39.50. [ISBN: 0-12-313340-8] Presents a number of iterative methods for solving large, sparse systems of linear algebraic equations including both computational and theoretical aspects. Also includes a discussion of the use of iterative methods in the solution of multidimensional boundary-value problems. AO

Numerical Analysis, S(16-18), P. Implementation of Finite Element Methods for Navier-Stokes Equations. Francois Thomasset. Ser. in Computational Physics. Springer-Verlag, 1981, vii + 159 pp, $28. [ISBN: 0-387-10771-1] A concise introduction to the use of finite element methods in fluid mechanics. The emphasis is on implementation issues rather than on the underlying mathematics. AO

Numerical Analysis, P. Numerical Integration. Ed: G. Hämmerlin. ISNM 57. Birkhauser Boston, 1982, 275 pp, $29.95. [ISBN: 3-7643-1254-8] The proceedings of a conference held in Oberwolfach in October, 1981. Among the topics discussed in the 25 papers are the theory of quadrature, multidimensional numerical integration schemes, and the quadrature of functions with special properties (e.g., convexity). AO

Numerical Analysis, P. Méthodes de lagrangien augmenté. M. Fortin, R. Glowinski. Dunod, 1982, xvi + 320 pp, 170 FF. [ISBN: 2-04-015465-5] The object of the book is to show that augmented Lagrange methods provide highly effective algorithms for the numerical solution of boundary value problems for partial differential equations. LCL

Numerical Analysis, T(17-18: 1), P. Finite Elements: Mathematical Aspects, Volume IV. J. Tinsley Oden, Graham F. Carey. Prentice-Hall, 1983, viii + 195 pp, $29. [ISBN: 0-13-317081-0] This work presents the basic mathematics needed to understand the convergence and stability properties of finite element methods for linear elliptic boundary-value problems. Some experience with finite

element methods and some knowledge of real and functional analysis is presumed. AO

Numerical Analysis, P. Lecture Notes in Mathematics-953: Iterative Solution of Nonlinear Systems of Equations. Ed: R. Ansorge, Th. Meis, W. Törnig. Springer-Verlag, 1982, vii + 202 pp, $12 (P). [ISBN: 0-387-11602-8] Proceedings representing 13 of 24 lectures presented at a meeting held at Oberwolfach, Germany, January 31 to February 5, 1982. JAS

Numerical Analysis, P. Numerical Grid Generation. Ed: Joe F. Thompson. Elsevier North-Holland, 1982, xxv + 909 pp, $95. [ISBN: 0-444-00757-1] Proceedings of a symposium on the numerical generation of curvilinear coordinate systems and their use in the numerical solution of partial differential equations, held in Nashville in April, 1982. 45 papers, many of which are expository. RWN

Numerical Analysis, S(13-15). Numerical Analysis Demonstrations on the HP-33E. Peter and Marie Louise Henrici. Wiley, 1982, 234 pp, $11.95 (P). [ISBN: 0-471-05943-9] A machine-dependent supplement to the first author's Essentials of Numerical Analysis, With Pocket Calculator Demonstrations (TR, January 1983). Nonetheless, an interesting, useful set of examples. RWN

Numerical Analysis, P. Nonlinear Equivalence, Reduction of PDEs to ODEs and Fast Convergent Numerical Methods. E.E. Rosinger. Research Notes in Math., No. 77. Pitman Pub, 247 pp, $21.95 (P). [ISBN: 0-273-08570-0] Integrates several of the author's works on several numerical methods for solving nonlinear systems of evolution partial differential equations. RWN

Functional Analysis, P. From A to Z: Proceedings of a Symposium in Honour of A.C. Zaanen. Ed: C.B. Huijsmans, et al. Math. Centre Tracts, No. 149. Math Centrum, 1982, vii + 130 pp, Dfl. 16,80 (P). [ISBN: 90-6196-241-2] Collection of papers by former Ph.D. students of Zaanen, primarily in functional analysis, integration theory and Riesz space theory. JRG

Functional Analysis, P. Schrödinger-type Operators with Continuous Spectra. M.S.P. Eastham, H. Kalf. Research Notes in Math., No. 65. Pitman Pub, 1982, 281 pp, $24.95 (P). [ISBN: 0-273-08526-3] An account of recent results concerning the existence and non-existence of eigenvalues embedded in the continuous spectrum of Sturm-Liouville and Schrödinger-type operators. AO

Functional Analysis, P. Spectral Theory and Wave Operators for the Schrödinger Equation. A.M. Berthier. Research Notes in Math., No. 71. Pitman Pub, 1982, 306 pp, $24.95 (P). [ISBN: 0-273-08562-X] This work gives a nearly complete spectral analysis of self-adjoint Schrödinger operators with short range potentials. AO

Functional Analysis, T(18: 1, 2), S, P. Lecture Notes in Mathematics-918: Schauder Bases in Banach Spaces of Continuous Functions. Zbigniew Semadeni. Springer-Verlag, 1982, v + 136 pp, $9.80 (P). [ISBN: 0-387-11481-5] The objects of study are countable subsets whose linear span is dense. General properties of Schauder bases of C(X), construction of spline and other special bases, and nonexistence results for certain classes of bases are developed. In textbook style, with exercises. PZ

Functional Analysis, P. Toeplitz Centennial. Ed: I. Gohberg. Operator Theory, V. 4. Birkhauser Verlag, 1982, 588 pp, $48.40. [ISBN: 3-7643-1331-1] Proceedings of a May 1981 memorial conference in Tel Aviv on the centennial of Otto Toeplitz' birth: 29 research papers on aspects of operator theory, plus four memorial papers (by Köthe, Dieudonne, and Toeplitz' son) describing Toeplitz' mathematical work as well as the personal and social pressures on a leading Jewish scholar during the rise of the National Socialists. LAS

Functional Analysis, P. Lecture Notes in Mathematics-948: Functional Analysis. Ed: D. Butković, H. Kraljević, S. Kurepa. Springer-Verlag, 1982, x + 239 pp, $12 (P). [ISBN: 0-387-11594-3] Lecture notes (excluding those of P.R. Halmos which have been published elsewhere) from a November 1981 conference at Dubrovnik, Yugoslavia. LAS

Functional Analysis, T(17-18: 1), S, P, L. Notes on Real and Complex C*-Algebras. K.R. Goodearl. Birkhauser Boston, 1982, 211 pp, $28.95. [ISBN: 0-906812-16-X] Intended as a "leisurely" introduction to C*-algebras, three major results are developed: the complex and real Gelfand-Naimark theory for both the commutative and non-commutative cases and the classification of approximately finite-dimensional C*-algebras. Exercises, references, index. JS

Functional Analysis, T(17-18: 1), S, P, L. A Short Introduction to Perturbation Theory for Linear Operators. Tosio Kato. Springer-Verlag, 1982, xiii + 161 pp, $19.80. [ISBN: 0-387-90666-5] With relatively minor changes and a modest expansion, this is essentially the first two chapters of the author's longer work on the subject (Perturbation Theory for Linear Operators, Springer-Verlag, 1966; TR's, August-September 1967; January 1978). It deals with the finite-dimensional theory and, as such, is considerably simpler and self-contained. JS

Functional Analysis, P. "Ultra"-Techniques in Banach Space Theory. Brailey Sims. Papers in Pure & Appl. Math., No. 60. Queen's U, 1982, v + 117 pp, (P). The Banach space ultraproduct has led to new developments and improved understanding in a variety of areas in Banach space theory. Written for Banach space theorists; assumes no prior knowledge of ultrapowers, and makes no explicit use of model theory. LCL

Analysis, P. Lecture Notes in Mathematics-919: Séminaire Pierre Lelong-Henri Skoda (Analyse) Années 1980/81 et Colloque de Wimereux, Mai 1981. Ed: Pierre Lelong, Henri Skoda. Springer-Verlag, 1982, vii + 386 pp, $17.40 (P). [ISBN: 0-387-11482-3]

Analysis, P. Lecture Notes in Mathematics-906: Séminaire de Théorie du Potentiel Paris, No. 6. F. Hirsch, G. Mokobodzki. Springer-Verlag, 1982, iv + 328 pp, $15.40 (P). [ISBN: 0-387-11185-9]

Analysis, P. Lecture Notes in Mathematics-939: Martingale Theory in Harmonic Analysis and Banach Spaces. Ed: J.-A. Chao, W.A. Woyczyński. Springer-Verlag, 1982, viii + 225 pp, $12 (P). [ISBN: 0-387-11569-2] The proceedings, less ten lectures by Donald L. Burkholder, of the NSF-CBMS conference held at the Cleveland State University, Cleveland, Ohio, July 13-17, 1981. JAS

Analysis, P, L.** Handbook of Applicable Mathematics, Volume IV: Analysis. Ed: Walter Ledermann, Steven Vajda. Wiley, 1982, xxiii + 865 pp, $85. [ISBN: 0-471-10141-9] Fourth of six volumes on core mathematics designed to set forth mathematical tools for practitioners in a self-contained, direct manner. This volume covers calculus, real and complex analysis, ordinary, partial and stochastic differential equations, integral transforms, nonlinear optimization, functional analysis, and much more. LAS

Analysis, P. Nonlinear Analysis and Applications. Ed: S.P. Singh, J.H. Burry. Lect. Notes in Pure & Appl. Math., V. 80. Dekker, 1982, x + 468 pp, $49.75 (P). [ISBN: 0-8247-1790-2] The proceedings of an international conference held at Memorial University, St. John's, Newfoundland, Canada from June 1-3, 1981. JAS

Analysis, P, L. Conference on Harmonic Analysis in Honor of Antoni Zygmund. Ed: William Beckner, et al. Wadsworth Pub, 1983, $79.75 set [ISBN: 0-534-98043-0]. Volume I, xvi + 345 pp; Volume II, xvi + 488 pp. Papers on trigonometric series, Fourier analysis, singular integrals, Hardy spaces, differentiation theory, partial differential equations, and harmonic analysis, from a March 1981 conference at the University of Chicago on the occasion of Zygmund's eightieth birthday. LAS

Algebraic Geometry, P. Lecture Notes in Mathematics-947: Algebraic Threefolds. Ed: Alberto Conte. Springer-Verlag, 1982, vii + 315 pp, $16.50 (P). [ISBN: 0-387-11587-0] Proceedings of the CIME conference held at the "Villa Monastero" in Varenna (Como), Italy, June 14-23, 1981. JAS

Algebraic Geometry, P. Lecture Notes in Mathematics-943: Modifications Analytiques. Vincenzo Ancona, Giuseppe Tomassini. Springer-Verlag, 1982, 120 pp, $8 (P). [ISBN: 0-387-11570-6]

Algebraic Geometry, P. Introduction to Algebraic and Abelian Functions, Second Edition. Serge Lang. Grad. Texts in Math., No. 89. Springer-Verlag, 1982, ix + 169 pp, $32. [ISBN: 0-387-90710-6] Among the changes and additions from the First Edition (TR, January 1973) are Weil's proof of Riemann-Roch, Rohrlich's work on the Fermat curve, and an expanded discussion of theta functions. SG

Differential Geometry, T(18), S, P. Singularities of Smooth Functions and Maps. Jean Martinet. Transl: Carl P. Simon. London Math. Soc. Lect. Note Ser., No. 58. Cambridge U Pr, 1982, xiv + 256 pp, $19.95 (P). [ISBN: 0-521-23398-4] A geometric treatment using only the basic notions of algebra (modules, ideals), topology (function spaces), and analysis (Taylor's theorem, implicit function theorem). The emphasis is on the local theory and the important role of unfolding and deformation. Among ideas discussed in detail are Morse theory, the division theorem for smooth functions, and several classification theorems, including Thom's theory of the seven elementary catastrophes. LCL

Differential Geometry, P. Lecture Notes in Mathematics-949: Harmonic Maps. Ed: R.J. Knill, M. Kalka, H.C.J. Sealey. Springer-Verlag, 1982, 158 pp, $10 (P). [ISBN: 0-387-11595-1] Contributed papers from the December 1980 CBMS Regional Conference at Tulane, intended as a companion volume to the CBMS volume (Springer Lecture Note No. 950) containing the principal lectures of James Eells. LAS

Differential Geometry, P. Selected Papers of Kentaro Yano. Ed: Morio Obata. Math. Stud., No. 70. Elsevier North-Holland, 1982, liii + 145 pp, $58.25 (P). [ISBN: 0-444-86495-4] A selection (by Yano) of his research papers in differential geometry, introduced by S.S. Chern and by a mathematical autobiography by Yano. LAS

Geometry, T*(15-16: 1, 2), S, L.** The Non-Euclidean, Hyperbolic Plane: Its Structure and Consistency. Paul Kelly, Gordon Matthews. Universitext. Springer-Verlag, 1981, xiii + 333 pp, $24 (P). [ISBN: 0-387-90552-9] With a detailed chapter on absolute geometry serving as foundation, authors present a careful and thorough--yet elementary--treatment of hyperbolic geometry. Use of transformations contributed novelty, efficiency and elegance. The Poincaré model, which is used to establish relative consistency, is studied in great detail. Many exercises and an appendix with brief treatment of other geometries. Excellent textbook for prospective high school teachers. SS

Geometry, P. Lecture Notes in Mathematics-926: Geometric Techniques in Gauge Theories. Ed: R. Martini, E.M. de Jager. Springer-Verlag, 1982, ix + 219 pp, $12.50 (P). [ISBN: 0-387-11497-1] Proceedings of the Fifth Scheveningen Conference on Differential Equations, The Netherlands, August 23-28, 1981. JAS

Algebraic Topology, P. Etale Homotopy of Simplicial Schemes. Eric M. Friedlander. Annals of Math. Stud., No. 104. Princeton U Pr, 1982, vii + 190 pp, $26.50; $11.50 (P). [ISBN: 0-691-08288-X] An

exposition of the current state of étale homotopy theory, with applications to algebraic topology, algebraic geometry, and the cohomology of groups. SG

Topology, P. Topologia. J. Margalef Roig, E. Outerelo Dominguez, J.L. Pinilla Ferrando. Editorial Alhambra, 1982, ix + 382 pp, (P). [ISBN: 84-205-0879-9] Chapters 12 and 13 of an encyclopedia of topology. Covers connectedness and spaces of functions. In Spanish. KS

Operations Research, S(15-17), L. Systems Modelling and Optimization. Ed: Peter Nash. Control Eng. Ser., V. 16. Peter Peregrinus, 1981, xii + 201 pp, $43 (P). [ISBN: 0-906048-63-X] The articles in this volume are drawn from material presented at a summer school for control engineers in July, 1980. Approximately one-half of the book is an introductory survey of optimization theory. The remainder is made up of five case studies and a short chapter on model evaluation. AO

Probability, P. Lecture Notes in Mathematics-921: Séminaire de Probabilités XVI, 1980/81 Supplément: Géométrie Différentielle Stochastique. Ed: J. Azéma, M. Yor. Springer-Verlag, 1982, 285 pp, $12.90 (P). [ISBN: 0-387-11486-6] A supplement to Lecture Notes 920 which covers the special lectures on stochastic differential geometry. JAS

Probability, P. Lecture Notes in Mathematics-928: Probability Measures on Groups. Ed: H. Heyer. Springer-Verlag, 1982, x + 477 pp, $23 (P). [ISBN: 0-387-11501-3] Proceedings of the Sixth Conference held at Oberwolfach, Germany, June 28 to July 4, 1981. JAS

Statistics, T(16-18: 1, 2), P*. Bayesian Reliability Analysis. Harry F. Martz, Ray A. Waller. Wiley, 1982, xix + 745 pp, $44.95. [ISBN: 0-471-86425-0] In the Wiley Series in Probability and Mathematical Statistics. Comprehensive, self-contained presentation. First quarter provides an introduction to probability, statistics, and classical reliability. Last three quarters deals exclusively with Bayesian methods for reliability analysis. Good sets of references. RSK

Statistics, P. Rates of Convergence in the Central Limit Theorem. Peter Hall. Research Notes in Math., N. 62. Pitman Pub, 1982, 251 pp, $22 (P). [ISBN: 0-273-08565-4] Research monograph unifying two disparate traditional approaches: upper bounds on rates of convergence, and characterizations of the rate of convergence. Presumes background in measure theory and Fourier analysis. RSK

Statistics, P. Sampling and Statistics Handbook for Research. Chester H. McCall, Jr. Iowa St U Pr, 1982, xi + 340 pp, $25.95 (P). [ISBN: 0-8138-1628-9] Introduction to basic statistical techniques and sampling theory for researchers in education. More concerned with sample surveys than experiments, it has chapters on stratified, cluster, and combination sampling plans. No exercises. RSK

Statistics, P. Edgeworth Expansions for Linear Combinations of Order Statistics. R. Helmers. Math. Centre Tracts, No. 105. Math Centrum, 1982, iv + 137 pp, Dfl. 17,85 (P). [ISBN: 90-6196-174-2] Technical monograph providing more precise estimates of linear combinations of order statistics which have limiting standard normal distributions. RSK

Computer Literacy, S(13). CBM Professional Computer Guide. Adam Osborne, Jim Strasma, Ellen Strasma. Osborne/McGraw-Hill, 1982, xi + 512 pp, $15 (P). [ISBN: 0-931988-75-6] This book is simply an extended user's manual for the Commodore CBM 8032 computer system. The entire book is highly specific for the hardware and software of that particular system. For teaching purposes it would probably need to be supplemented by an introductory programming text and/or an introductory text on microcomputers. For its intended audience, however (owners of CBM 8032 systems), it seems quite thorough. It takes them through CBM-Basic, screen editing, peripherals, mass storage and advanced features of the hardware. This would be a useful book to purchase at the time that you buy a CBM 8032 personal computer. MS

Computer Programming, S(13), P. Program Your Microcomputer in BASIC. P.E. Gosling. Dilithium Pr, 1981, viii + 91 pp, $5.95 (P). [ISBN: 0-918398-52-5] An introduction to Basic primarily for microcomputers. The text consists largely of examples of programs. CEC

Computer Programming, T*(13: 1), L. An Introduction to Programming and Problem Solving with Pascal, Second Edition. G. Michael Schneider. Wiley, 1982, xi + 468 pp, $21.95. [ISBN: 0-471-08216-3] This textbook is designed for a first course in computer programming similar to the course CS1 outlined in the ACM Curriculum '78 report. It covers all aspects of programming from problem specification through documentation and maintenance. AO

Computer Programming, S. Practical BASIC Programs: IBM Personal Computer Edition. Ed: Lon Poole. Osborne/McGraw-Hill, 1982, x + 162 pp, $15.99 (P). [ISBN: 0-931988-80-2] Collection of forty programs from areas of finance, management decision making, statistics, and mathematics. Each program includes brief introduction to subject matter, description of input and output, sample run, practice problems, and complete listing. KS

Computer Programming, T(13: 1), S. Learning BASIC Step by Step, Teacher's Guide. Vern McDermott, Diana Fisher. Computer Sci Pr, 1982, vii + 257 pp, $17.95 (P). [ISBN: 0-914894-33-1] Another Basic programming text. Introduces elements of language through demonstration problems. Extensive use of flowcharts. Teacher's guide contains more detailed explanations, solutions, comments on common errors and other pedagogical hints. KS

Computer Programming, T*(13: 1), L. Advanced Programming and Problem Solving with Pascal. G. Michael Schneider, Steven C. Bruell. Wiley, 1981, xiii + 506 pp, $26.95. [ISBN: 0-471-07876-X] A textbook for a second course in computer programming similar to the course CS2 of the ACM Curriculum ´78 report. A special feature of this book is the inclusion of relatively large, complex case studies to illustrate key points. AO

Computer Programming. Timex Sinclair 1000: Programs, Games and Graphics. Robin Jones, Ian Stewart. Birkhauser Boston, 1982, vii + 156 pp, $10.95 (P). [ISBN: 3-7643-3080-5] A non-threatening, witty but yet sound introduction to the Timex-Sinclair 1000. Includes sections on both hardware and Basic along with several program listings of applications. CEC

Software Systems, S(16-18), P. Lecture Notes in Computer Science-143: Operating Systems Engineering. Ed: M. Maekawa, L.A. Belady. Springer-Verlag, 1982, vii + 465 pp, $19 (P). [ISBN: 0-387-11604-4] The proceedings of an IBM-sponsored conference held in Amagi, Japan, in October, 1980. The papers present recent work on the engineering of operating systems organized under the headings: concurrency and access control, program behavior and performance models, operating system evaluation, user interfaces, distributed operating systems, network operating systems, development processes and tools, and data flow machines. AO

Computer Science, P, L. The ILLIAC IV: The First Supercomputer. R. Michael Hord. Computer Sci Pr, 1982, xii + 350 pp, $29.95. [ISBN: 0-914894-71-4] The Illiac IV (operational in 1975) was the first large scale array computer, incorporating both pipelining and a high level of parallelism. Intended for computer professionals, this book details the history, structure, applications, and impact of the Illiac IV, providing a perspective on its strengths and weaknesses. GHM

Computer Science, S(15-17), L. Computer Architecture: A Structured Approach. R.W. Doran. APIC Studies in Data Processing, No. 15. Academic Pr, 1979, xi + 233 pp, $42. [ISBN: 0-12-220850-1] An introduction to the general principles of computer architecture and a detailed examination of their implementation in the Burroughs B6700 and B7700 computers. AO

Computer Science, T(14-15: 1, 2). Mathematical Structures for Computer Science. Judith L. Gersting. WH Freeman, 1982, xi + 452 pp, $22.50. [ISBN: 0-7167-1305-5] A broad overview of some basic structures and techniques involving logic, sets, graphs, Boolean algebras, groups, morphisms, simple coding, finite state and Turing machines, and formal languages. Informal style with many applications, but also an appreciable introduction to the rigorous theoretical framework for expressing ideas in computer science. GHM

Computer Science, P. An Analysis of Sparse Matrix Storage Schemes. M. Veldhorst. Math. Centra Tracts, No. 150. Math Centrum, 1982, vii + 237 pp, Dfl. 30,45 (P). [ISBN: 90-6196-242-0] A number of storage schemes for sparse matrices are proposed and algorithms for manipulating them are analyzed in terms of their computational complexity. AO

Computer Science, P. Lecture Notes in Computer Science-142: Problems and Methodologies in Mathematical Software Production. Ed: P.C. Messina, A. Murli. Springer-Verlag, 1982, 271 pp, $12.50 (P). [ISBN: 0-387-11603-6] The presentations from an international seminar held at Sorrento, Italy, November 3-8, 1980. JAS

Computer Science, T(14-15), S*. Data Base Management Systems: A Guide to Microcomputer Software. David Kruglinski. Osborne/McGraw-Hill, 1983, ix + 260 pp, $16.95 (P). [ISBN: 0-931988-84-5] Three chapters of elementary but substantial theory are followed by detailed discussions of four popular microcomputer systems: Condor, dBASE II, FMS-80, and MDBS III. The book closes with some remarks about the future, other microcomputer-based data base systems, and some sample programs for dBASE II and FMS-80. Its biggest weakness as an elementary text is the lack of exercises. A well-written approach that provides substantial insight about what´s going on--much more than a user´s manual. JAS

Computer Science, P, L*. Campus Computing Strategies. Ed: John W. McCredie. Digital Pr, 1983, xi + 316 pp, $21. [ISBN: 0-932376-20-7] 10 chapters by directors of college and university computer centers providing case studies in organization, planning and management of campus computer services. The book evolved from a 1981-82 study by EDUCOM, a national consortium specializing in academic computer services. LAS

Control Theory, P. Contrôle impulsionnel et inéquations quasi variationnelles. A. Bensoussan, J.L. Lions. Dunod, 1982, xv + 596 pp, 280 FF. [ISBN: 2-04-015428-0] Uses dynamic programming to establish a relationship between stochastic control and quasi-variational inequalities in a way that leads to numerical applications. LAS

Control Theory, T(18), P. Stochastic Control by Functional Analysis Methods. Alain Bensoussan. Stud. in Math. & Its Appl., V. 11. Elsevier North-Holland, 1982, xv + 410 pp, $58.25. [ISBN: 0-444-86329-X] An advanced textbook on stochastic control, containing basic results, relying on functional analysis and the theory of partial differential equations. LCL

Systems Theory, T(17-18: 1, 2), P. System Theory: A Hilbert Space Approach. Avraham Feintuch, Richard Saeks. Pure & Appl. Math., V. 102. Academic Pr, 1982, xiii + 310 pp. [ISBN: 0-12-251750-4] A unified theory of linear systems using Hilbert space techniques. Also presents the theory of operators defined on a Hilbert resolution space (essentially a Hilbert space with a time structure

axiomatically adjoined). AO

Applications, T(16-17: 1, 2), L. Applied Mathematics: Principles, Techniques, and Applications. James A. Cochran. Wadsworth Pub, 1982, x + 399 pp, $34.95. [ISBN: 0-534-98026-0] A textbook for an advanced course in modern applied mathematics. The topics covered include the calculus of variations, stability theory, conformal mappings, generalized functions, integral equations, and asymptotics. AO

Applications, P. Nonlinear Problems: Present and Future. Ed: Alan Bishop, David Campbell, Basil Nicolaenko. Math. Stud., No. 61. Elsevier North-Holland, 1982, xi + 483 pp, $69.75 (P). [ISBN: 0-444-86395-8] Proceedings of an interdisciplinary conference inaugurating the Los Alamos Center for Nonlinear Studies: turbulence, reaction-diffusion processes, nonlinear analysis, and low dimensional solids. LAS

Applications, S, P, L*. Iterated Maps on the Interval as Dynamical Systems. Pierre Collet, Jean-Pierre Eckmann. Progress in Physics, No. 1. Birkhauser Boston, 1980, vii + 248 pp, $17.50. [ISBN: 3-7643-3026-0] An exposition for mathematicians and theoretical physicists of the fascinating (and often perplexing) connection among iterated continuous maps of the interval, dissipative dynamical systems (e.g., physical systems with friction) when driven by an outside force, and "chaotic," "turbulent" or "strange" phenomena in biological, chemical, physical and economic systems. LAS

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-202: SMS--A Program Package for Simulation and Gaming of Stochastic Market Processes and Learning Behavior. Ulrich Witt, Joachim Perske. Springer-Verlag, 1982, vii + 266 pp, $18.50 (P). [ISBN: 0-387-11551-X] SMS is a package of Fortran programs for Stochastic Market Simulation. This work describes the use of SMS for experimental programming and Monte Carlo simulation studies. AO

Applications (Economics), P. Philosophy of Economics. Ed: W. Stegmüller, W. Balzer, W. Spohn. Stud. in Contemp. Econ. Springer-Verlag, 1982, viii + 306 pp, $21 (P). [ISBN: 0-387-11927-2] Proceedings of a 1981 conference in Munich, the beginnings of a new philosophical enterprise. In three parts: structuralist theories (the use of model-theoretic tools in the analysis of theories); applications of structuralist theories to Marxian theory; and decision theory as a basis for economic activity. LAS

Applications (Economics), P, L. Lecture Notes in Economics and Mathematical Systems-207: Modern Analysis of Value Theory. Y. Fujimori. Springer-Verlag, 1982, 165 pp, $12 (P). [ISBN: 0-387-11949-3] An exposition in terms of Leontief and von Neumann models of Marx's theory of value that explains economic relationships among goods in terms of the value of human labor which is assumed invariant. LAS

Applications (Economics), P. Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Robert J. Aumann, et al. Bibliographisches Institut, 1981, 196 pp, 42 DM (P). [ISBN: 3-411-01609-4] Nine papers presented at a June 1979 memorial symposium at the University of Mannheim. LAS

Applications (Fluid Flow), T(17-18: 1), S. Computational Methods for Fluid Flow. Roger Peyret, Thomas D. Taylor. Ser. in Computational Physics. Springer-Verlag, 1983, x + 358 pp, $42.50. [ISBN: 0-387-11147-6] Numerical methods for solving Navier-Stokes equations, including finite difference, finite element, spectral and special methods. Applied to incompressible and inviscid and viscid compressible flows. A useful introduction to this subject. RWN

Applications (Game Theory), T(16-18), S, P*, L. Game Theory in the Social Sciences: Concepts and Solutions. Martin Shubik. MIT Pr, 1982, ix + 514 pp, $35. [ISBN: 0-262-19195-4] A pluralistic use of n-person game theory to reshape the methodology of theoretical economics: different concepts of "solution" reflect various economic models in ad hoc yet pragmatic arguments. Begins with several chapters devoted to mathematical modelling in the social sciences; then introduces the variety of solution concepts provided by the theory of games. A second volume will be devoted to specific economic applications. LAS

Applications (Humanities), P*, L. Computing in the Humanities. Ed: Richard W. Bailey. Elsevier North-Holland, 1981, viii + 191 pp, $46.50. [ISBN: 0-444-86423-7] Papers selected from the Fifth International Conference on Computers in the Humanities held in May 1981 at Ann Arbor. Topics range from morphology to typography (TEX for humanists), from remedial teaching of foreign languages and UNIX aids for composition courses to data bases for biblical texts. "The extension of humanistic inquiry through data processing is merely scholarship carried out by other means." LAS

Applications (Management), S, P*, L.** The Art and Science of Negotiation. Howard Raiffa. Harvard U Pr, 1982, x + 373 pp, $18.50. [ISBN: 0-674-04812-1] Case studies in negotation--political, civil, economics--ranging from the fair division of estates to the law of the sea. Organized along game-theoretic lines (two parties, one issue; two parties, many issues; several parties, many issues), the analysis of cases employs an informal mix of elementary algebra, game theory, linear programming, and other heuristic optimization techniques. LAS

Applications (Modelling), T*(14-15: 1), S, L. Applying Mathematics: A Course in Mathematical Modelling. David Burghes, Ian Huntley, John McDonald. Math. & Its Applic. Ellis Horwood, 1982, 194 pp, $44.95. [ISBN: 0-85312-417-5] Appealing yet simple models (aircraft separation, broad jump at Mexico

City, communication satellites) introduce realistic modelling situations. Brief discussion of strategies and philosophy of modelling then leads to a rich collection of suggestions for modelling projects. An excellent resource for a modelling seminar, using little more than elementary calculus. LAS

Applications (Modelling). Air Pollution: Assessment Methodology and Modeling. Ed: Eric Weber. NATO: Challenges of Mod. Soc., V. 2. Plenum Pr, 1982, viii + 329 pp, $39.50. [ISBN: 0-306-40997-6] A comprehensive, state-of-the-art survey of air quality control modelling prepared for NATO as part of a German pilot study, in cooperation with Belgium and the U.S. EPA. Concludes with a 60 page glossary of modelling and air quality control terminology. LAS

Applications (Physics), L. Formulas, Facts and Constants for Students and Professionals in Engineering, Chemistry and Physics. H.J. and K.H. Fischbeck. Springer-Verlag, 1982, xii + 251 pp, $14 (P). [ISBN: 0-387-11315-0] Includes mathematical formulae (tables of integrals, etc.), conversion factors and values of fundamental constants for the SI system, the basic terms of spectroscopy, atomic structure, and wave mechanics, and physical data of use in the laboratory. AO

Applications (Physics), T(16-18: 1), P*, L. Quantum Mechanics of Atoms and Molecules. Walter Thirring. Course in Math. Physics, No. 3. Transl: Evans M. Harrell. Springer-Verlag, 1981, viii + 300 pp, $28. [ISBN: 0-387-81620-8] An axiomatic introduction to quantum mechanics which emphasizes concrete results that can be compared with experimental data. A good illustration of the use of modern functional analysis in mathematical physics. AO

Applications (Physics), P. Quantum Mechanics in Mathematics, Chemistry, and Physics. Ed: Karl E. Gustafson, William P. Reinhardt. Plenum Pr, 1981, ix + 506 pp, $59.50. [ISBN: 0-306-40737-X] The 37 papers in this volume were originally presented at the March, 1980 AMS meeting in Boulder, Colorado at a special session on mathematical physics. AO

Applications (Physics), P. Iterative Methods for Calculating Static Fields and Wave Scattering by Small Bodies. Alexander G. Ramm. Springer-Verlag, 1982, xii + 122 pp, $24 (P). [ISBN: 0-387-90682-7] This monograph presents iterative methods for solving exterior and interior static boundary-value problems. Most of the problems discussed are three-dimensional. AO

Applications (Physics), S(17-18), P. Spacetime and Geometry: The Alfred Schild Lectures. Ed: Richard A. Matzner, L.C. Shepley. U of Texas Pr, 1982, x + 189 pp, $37.50. [ISBN: 0-292-77567-9] Seven expository lectures given over a period of several years starting in 1977. The lectures presented here are: "Why Is the Universe So Symmetrical?" by Dennis Sciama; "Null Congruences and Plebanski-Schild Spaces," by Ivor Robinson; "Linearization Stability," by Dieter Brill; "Nonlinear Model Field Theories Based on Harmonic Mappings," by Charles W. Misner; "Gravitational Fields in General Relativity," by Roy P. Kerr; "On the Potential Barriers Surrounding the Schwarzschild Black Hole," by Subrahmanyan Chandrasekhar; "The Initial Value Problem and Beyond," by James W. York, Jr. and Tsvi Piran. JAS

Applications (Physics), P. Lecture Notes in Physics-162: Relativistic Action at a Distance: Classical and Quantum Aspects. Ed: J. Llosa. Springer-Verlag, 1982, x + 263 pp, $15 (P). [ISBN: 0-387-11573-0] Proceedings of a workshop held in Barcelona, Spain, June 15-21, 1981. JAS

Applications (Physics), P. General Relativity: An Introduction to the Theory of the Gravitational Field. Hans Stephani. Ed: John Stewart. Transl: Martin Pollock, John Stewart. Cambridge U Pr, 1982, xvi + 298 pp, $49.50. [ISBN: 0-521-24008-5] An introduction to the general theory of relativity for readers familiar with theoretical mechanics, electrodynamics, and special relativity. The essential concepts and formulae of Riemannian geometry are presented in the first few chapters. AO

Applications (Physics), T. Introduction to Special Relativity. Wolfgang Rindler. Clarendon Pr, 1982, x + 185 pp, $13.95 (P); $39. [ISBN: 0-19-853182-6; 0-19-853181-8] A nicely written text. Lots of exercises and a good index. Tensors are taken classically and the author has no qualms about working in three-space rather than four-space when it's advantageous to clarifying his point. However, the essence is space-time geometry and tensors. JAS

Applications (Physics), T(117-18: 1, 2), P*. Classical Fields: General Relativity and Gauge Theory. Moshe Carmeli. Wiley, 1982, xvii + 650 pp, $44.95. [ISBN: 0-471-86437-4] This text presents a unified treatment of gauge fields and gravitational fields. Approximately one-half of the work is devoted to each topic. AO

Applications (Social Science), P, L. Power, Voting, and Voting Power. Ed: Manfred J. Holler. Physica-Verlag, 1982, 338 pp, $69.95 (P). [ISBN: 3-7908-0266-2] A normative and descriptive theory of democracy subsumed under two themes--theory of social choice and theory of games. A carefully planned, international volume intended to address the issues outlined in Arrow's 1951 Nobel-prize winning monograph on social choice. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St.

Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

New Jersey Section

The fall meeting of the New Jersey section was held on October 31, 1982 at Rutgers University's Newark campus. Thirty people registered for the meeting.

Invited Addresses:

"Time Warps, String Edits, and Macromolecules: The Mathematics of Sequence Comparisons," by Joseph Kruskal, Bell Labs.
"A Mathematician's London," by Dorothy Goldberg, Kean College.

Panel Discussion:

"On the Retraining of Mathematicians in Computer Science," by Susan Marchand, Kean College (Moderator); Robert Gillman, Stevens Institute of Technology; Anita La Salle and James Mc Hugh, New Jersey Institute of Technology; Richard Pirchner, Monmouth College.

Student Paper:

"Non-Associative Simple Rings," by Sin Min Lee, Stevens Institute of Technology.

Eastern Pennsylvania and Delaware Section

The annual fall meeting of the Eastern Pennsylvania and Delaware Section was held on November 20, 1982 at West Chester State College. There were approximately 125 persons in attendance.

Invited Lectures:

"Some Bijective Proofs in Combinatorics," by Herbert S. Wilf, University of Pennsylvania.
"Descartes, Euler, and Polyhedra," by Peter J. Hilton, State University of New York at Binghamton.
"Mathematical Science Curricula," by Alan C. Tucker, State University of New York at Stony Brook.

Panel Discussion:

"Mathematical Science Curricula," by Gerald J. Porter, University of Pennsylvania (Moderator); Jeorg Mayer, Lebanon Valley College, and Larry E. Sigler, Bucknell University.

Seaway Section

The Seaway Section held its fall meeting at St. Lawrence University, Canton, New York on November 12-13, 1982 with a registration of 61 persons.

Invited Address:

"Retraining Mathematicians to Teach Computer Science," by Ed Dubinsky, Clarkson College.

Papers Presented:

"What Makes a Ring Finite?" by Howard E. Bell, Brock University.
"Understanding and Teaching Mathematical Problem Solving," by Alan H. Schoenfeld, University of Rochester.
"A Follow-Up Study of the Winners of the U.S.A. Mathematical Olympiad," by Nura D. Turner, State University of New York at Albany.
"Graphical Solutions of Systems of n Linear Equations with n Equal to or Greater Than Three," by Robert D. Larsson, North Country Community College.
"Teaching Statistics with Computers," by Chanchal Singh, St. Lawrence University.
"Some Interesting Properties of the Integer Function," by Robert W. Sloan, Alfred University.

Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

New Jersey Section

The fall meeting of the New Jersey section was held on October 31, 1982 at Rutgers University's Newark campus. Thirty people registered for the meeting.

Invited Addresses:

"Time Warps, String Edits, and Macromolecules: The Mathematics of Sequence Comparisons," by Joseph Kruskal, Bell Labs.
"A Mathematician's London," by Dorothy Goldberg, Kean College.

Panel Discussion:

"On the Retraining of Mathematicians in Computer Science," by Susan Marchand, Kean College (Moderator); Robert Gillman, Stevens Institute of Technology; Anita La Salle and James Mc Hugh, New Jersey Institute of Technology; Richard Pirchner, Monmouth College.

Student Paper:

"Non-Associative Simple Rings," by Sin Min Lee, Stevens Institute of Technology.

Eastern Pennsylvania and Delaware Section

The annual fall meeting of the Eastern Pennsylvania and Delaware Section was held on November 20, 1982 at West Chester State College. There were approximately 125 persons in attendance.

Invited Lectures:

"Some Bijective Proofs in Combinatorics," by Herbert S. Wilf, University of Pennsylvania.
"Descartes, Euler, and Polyhedra," by Peter J. Hilton, State University of New York at Binghamton.
"Mathematical Science Curricula," by Alan C. Tucker, State University of New York at Stony Brook.

Panel Discussion:

"Mathematical Science Curricula," by Gerald J. Porter, University of Pennsylvania (Moderator); Jeorg Mayer, Lebanon Valley College, and Larry E. Sigler, Bucknell University.

Seaway Section

The Seaway Section held its fall meeting at St. Lawrence University, Canton, New York on November 12-13, 1982 with a registration of 61 persons.

Invited Address:

"Retraining Mathematicians to Teach Computer Science," by Ed Dubinsky, Clarkson College.

Papers Presented:

"What Makes a Ring Finite?" by Howard E. Bell, Brock University.
"Understanding and Teaching Mathematical Problem Solving," by Alan H. Schoenfeld, University of Rochester.
"A Follow-Up Study of the Winners of the U.S.A. Mathematical Olympiad," by Nura D. Turner, State University of New York at Albany.
"Graphical Solutions of Systems of n Linear Equations with n Equal to or Greater Than Three," by Robert D. Larsson, North Country Community College.
"Teaching Statistics with Computers," by Chanchal Singh, St. Lawrence University.
"Some Interesting Properties of the Integer Function," by Robert W. Sloan, Alfred University.

"A Simulation Approach to Evaluating Experimental Designs," by Robin H. Lock, Clarkson College.
"The Role of Monotone Method in the Study of Differential Equations," by Srinivasa Leela, State University College at Geneseo.

Northeastern Section

The annual meeting of the Northeastern Section was held at Worcester State College on November 19-20, 1982. Approximately 140 attended the meeting.

Invited Lectures:

* "CUPM's Recommendation for a Mathematics Curriculum in the 1980's," by Alan Tucker, SUNY at Stony Brook.
"How to Succeed in Business With and Without Mathematics," by Richard Franklin, Director of Corporate Development, Instrumentation Labs.
"The Maximum Principle and Analytic Functions (Christie Lecture)," by John Wermer, Brown University.
"A Guided Discovery Approach to Calculus," by Donald Small, Colby College.

Panel Discussions:

"Textbook and Curriculum Needs for the 1980's," Bodh Gulati (Moderator) Southern Connecticut State College; Gary Ostedt, Wiley Publishers; Peter Divine, McGraw-Hill Publishers; Gary Folvin, Allyn and Bacon Publishers.
"Why Math, or Why and How Should We Teach Mathematics in the 1980's," Carroll McMahon (Moderator), Bentley College; Richard Goller, Sanders Associates; Alan Suchat, Wellesley.

Contributed Papers:

* "IRA's, the Higher Mathematics of 2 + 2, and the Fundamental Theorem of Inflation," by Robert Pease, Jr., Senior Editor, Science.
"Stereographic Projection," by Ernest Manfred, U.S. Coast Guard Academy.
* "Using the Computer in the Classroom," by John D. McKenzie, Jr., Babson College.
"Mathematical Systems Analysis--A Course," by John Goulet, Colby College.

Maryland-District of Columbia-Virginia Section

The Maryland-District of Columbia-Virginia Section met on November 12-13, 1982 at Gallaudet College in Washington, D.C.

Invited Addresses:

"The Mathematical Sciences Curriculum K-12: What is Still Fundamental and What is Not?" by Marcia Sward, Associate Director of the Mathematical Association of America.
"Applications of Statistics in the Telecommunications Industry," by James Maher, Bell Telephone Labs.

Short Presentations:

"On the Zeros of Riemann's Continuous Non-differentiable Function," by Lee Whitt, Daniel H. Wagner Associates.
"Self-inverse Integer Matrices," by Robert Hanson, James Madison University.
"Measuring the Effectiveness of Computer Networks," by Robert Lewand, Goucher College.
"Jeeps Are Not Scalars," by Carmen Castells, student at The Johns Hopkins University.
"Algorithm for Singular Value Decomposition," by Hoy Booker, American University.
"Math Anxiety in Minorities," by Queen E. Wiggs, University of the District of Columbia.
"Markov Chains and Work Histories of the Disabled," by John Hennessey, Loyola College and the Social Security Administration.
"Group Theory and Crystallography," by George Mackiw, Loyola College.
"Buckled Polyhedra and Other Developable Surfaces," by Michael Goldberg.
"Optimization of Cancer Chemotherapy and Radiotherapy," by Paul Massell, U.S. Naval Academy.
"Ancient and Modern Mathematics in the St. John's College Program," by Samuel Kutler, St. John's College.
"Rotationally Invariant Bergman Measures and Subharmonic Functions," by Richard B. Tucker, Mary Baldwin College.
"A Computer Graphics Tutorial," by Stefan Shrier, Systems Planning Corporation.
"Helping Students Improve Their Mathematics Learning Skills," by Lewis Hirsch, Rutgers University.
"Security Algebras for Databases, etc.," by John Hays, Naval Research Laboratory.
"Graphical Analysis Methods Applied to Examining the Relationship of Ability and Achievement Variables to Successful Problem Solving," by Carolyn Maher, Rutgers University.

Four faces with two mathematically famous names. The classical book bearing those names was for a time co-extensive with the whole of its field. See p. 211.

If, for example, our integral involves the product of a logarithmic function and an algebraic function, we choose u to be the log part and, keeping criterion (1) in mind, the algebraic part is chosen for dv. By making this selection, in $\int v\,du$, the logarithmic character has disappeared by virtue of the differentiation. The antidifferentiation of the algebraic part results (most of the time) in an algebraic v. Consequently, the integral $\int v\,du$ is strictly algebraic in its integrand and hopefully easier than our original integral.

Similarly, if we have any combination of two of these types of functions in our original integral, we choose for u that type that appears first in LIATE and dv is whatever is left, that part that appears second in LIATE.

Rationale: Differentiating logs and inverse trig functions changes their character to algebraic, while algebraic functions run the "middle of the road;" differentiating them yields algebraic, while integration sometimes yields nonalgebraic, so we would probably differentiate them before integrating. Trig and exponential functions "bring up the rear" since their differentiation or antidifferentiation results in similar functions.

As an example, consider $\int x\ln(x)\,dx$. Using LIATE, we pick $u = \ln(x)$ and $dv = x\,dx$. We obtain $du = x^{-1}\,dx$ and $v = x^2/2$. Consequently,

$$\begin{aligned}\int x\ln(x)\,dx &= \frac{x^2}{2}\ln(x) - \int\frac{x}{2}\,dx\\ &= \frac{x^2}{2}\ln(x) - \frac{x^2}{4} + C.\end{aligned}$$

Students have liked using LIATE and find that choosing u and dv is much less frightening. The author has yet to see an integral done by integration by parts to which LIATE cannot be successfully applied.

To be able to use integration by parts, perhaps it is better LIATE than never.

Reference

1. Howard Anton, Calculus with Analytic Geometry, Wiley, New York, 1980.

MISCELLANEA

95. For the most part [Glaisher] accepted or rejected contributions on his own responsibility. The best editors have of course a wonderful power of judging the quality of a manuscript without attempting to read it; but the task becomes steadily more difficult as mathematics grows and specializes... A private journal, controlled autocratically by a kindly and discriminating editor, can serve many useful purposes, and particularly that of giving early encouragement to beginners, on whom societies, with councils and referees, are sometimes very severe.

—G. H. Hardy, "Dr. Glaisher and the 'Messenger of Mathematics,'" *Messenger of Mathematics*, 58 (1929) 159–160.

ANSWER TO PHOTOS ON PAGE 195.

Top: P. S. Alexandroff and H. Hopf (topology). Bottom: A. D. Alexandroff and E. Hopf (analysis, of two different but overlapping kinds).

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N IN4.*

AN OLLA-PODRIDA OF OPEN PROBLEMS, OFTEN ODDLY POSED

RICHARD K. GUY

Many problems are submitted to the MONTHLY, but a large proportion don't appear in print, for a variety of reasons. Some problems submitted to this section are brief and with no known connexions with the literature, so that they don't fit the format we've been using. But this isn't a good reason for not airing them.

Some problems may not be well enough posed, so that the real problem is to decide what is the problem. This may have several solutions, and the resulting problems may vary from the trivial to the impossible, so that, once existence and lack of uniqueness have been established, an intermediate problem may be to select the solution which is the best problem! For example,

José M. Bayod, University of Santander, Santander, Spain,

notes that in some parts of his country, a farmer owns a part of the mountains that surround his farm, according to the "flowing water law" which can be stated thus:

> A spot of rain lands at A on the mountain and flows downhill until it reaches a point B on the farm. Then the owner of point B in the valley also owns the point A on the mountain.

The main problem is to build a mathematical model for the problem: given an area in the valley, what part of the mountains is associated with it? A rough formulation would be: given a set of contour lines, is there a (unique) set of orthogonal trajectories (lines of greatest slope)? Joel Brenner warns about the many pitfalls. It may be necessary to know how the law is interpreted. A peak, or even a ridge, can be neglected as being of measure zero, but what about a plateau, whether horizontal or sloping? In practice, contour lines often have a discontinuous tangent at a stream bed, while the stream beds are likely to be of critical importance. Finally, contours and stream beds are often time-dependent. How can mathematicians and lawyers help one another? R. B. Kusner suggests that one may wish to assume that the contour lines are the level curves of a Morse function f (i.e., a smooth function with isolated nondegenerate critical points) since Morse functions are dense in the continuous functions and their complement in the smooth functions has measure zero. Unique orthogonal trajectories to the contour lines are generated by the gradient field Df.

Bayod further writes that there are remarkably few difficulties in real life; perhaps because so far neither mathematicians nor lawyers have been around! It was only curiosity that led him to formulate the problem, not the need to fill a legal gap. Yet the mountain often has pines, very good for wood, and is very valuable.

Robert B. Kusner, University of California, Berkeley, CA 94720

sends a less controversial problem on r-equilateral sets which still leaves something to the reader's imagination, however.

Suppose r is a positive real number and M is a metric space with distance d. Then a subset E of M is called **r-equilateral**, and we write $d(E) = r$, if $d(x, y) = r$ for every pair of distinct points x, y in E. Also define

$$e(r, M) = \max\{\#E | d(E) = r\}$$

where $\#E$ is the cardinality of the set E, and call $e(M)$, the maximum of $e(r, M)$ taken over all positive real numbers, the **equilateral dimension** of M.

General Problem. Compute the equilateral dimension for some familiar metric spaces.

The most familiar examples of all are the normed linear spaces $L_p(n)$ defined on $\mathbb{R}^n$ for integers $p \geqslant 1$ by the norm

$$\|x\|_p = \left(\sum_{1 \leqslant i \leqslant n} |x_i|^p \right)^{1/p}.$$

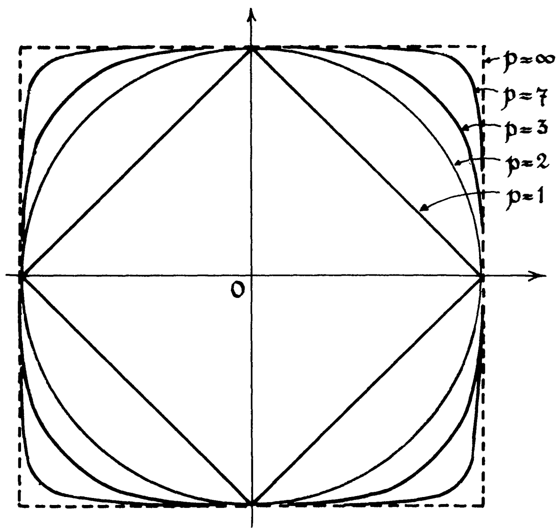


FIG. 1

Figure 1 is a picture of the unit "spheres" when $n = 2$. For $p = 1$ we have the inside square; $p = 2$ is a circle; $p = 3$ and $p = 7$ are also shown, and the dotted square is the limit as $p \to \infty$, the unit "sphere" in the space $L_\infty(n)$ with norm

$$\|x\|_\infty = \max_{1 \leqslant i \leqslant n} |x_i|.$$

If you look at pictures, you should be able to convince yourself that $e(L_p(n)) \geqslant n + 1$ for all p, $1 \leqslant p \leqslant \infty$, and all $n \geqslant 0$. On the other hand, you will find that $e(L_p(n)) \leqslant 2^n$, and Figure 2 may help you see that equality holds when $p = \infty$. When $p = 1$, the lower bound on $e(L_1(n))$ is $2n$.

Also the case $p = 2$ is quite special: every isometry (distance-preserving map) is affine, and all the affine maps generated by rotation and translation are isometries. Check that if $d(E) = r$ and $\#E = k + 1$, then the affine span of E has dimension k as an affine space. Therefore $e(L_2(n)) = n + 1$.

Problem 0. Does $e(L_1(n)) = 2n$?

Problem 1. Compute $e(L_p(n))$ for $2 < p < \infty$.

Kusner's guess is $n + 1$, but he notes that there are lots of numbers between $n + 1$ and 2^n. Here is a "solution sketch." In the range $1 < p < \infty$ the r-spheres are all smooth $(n - 1)$-

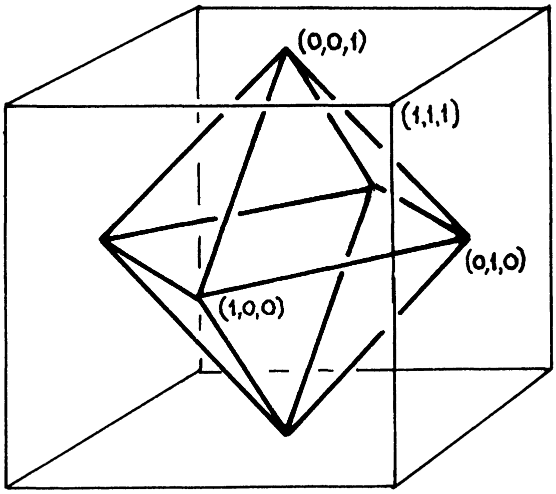


FIG. 2

manifolds. If $d(E) = r$ and $\#E = k$, then a point x such that $d(E, x) = r$ must lie on the k r-spheres about the points of E. Since the dimension of the intersection decreases by 1 "each time" ("general position") the biggest $\#E$ is $n + 1$. Certainly the nonsmoothness for $p = 1$ and $p = \infty$ account for the difference.

Riemannian manifolds are metric spaces which are "locally like $\mathbb{R}^n$." To see that $e(M) \geqslant n + 1$ for such an n-dimensional manifold M^n, let $x \in M^n$ and pick a point y on the (smooth) r-sphere about x. The r-sphere about y intersects the one about x in a set of dimension $\geqslant n - 2$. Now proceed inductively as in the above "solution sketch."

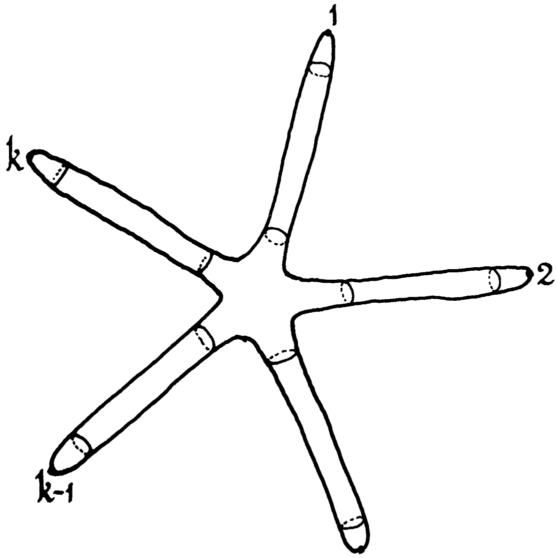


FIG. 3

Observe that $e(S^n) = n + 2$ for the round sphere $S^n \subset \mathbb{R}^{n+1} = L_2(n + 1)$. It may be that the following example shows that $e(M^2)$ is arbitrarily large. For any k, pick k points on a 2-dimensional (diffeomorphic but not isometric to the round) sphere M^2. Think of M^2 as made of

putty and pull out long fingers at these k points as in Figure 3. If the metric on M^2 is perturbed to correspond to this stretching, we get k points which are arbitrarily close to being equilateral. The open question is: can we go all the way to equilateral?

If *yes*, define $u(M)$ as the minimum of $e(M, r)$ taken over all positive reals. Then $u(M^2) = 3$ for these examples.

Problem 2. Compute an upper bound for $u(M^n)$ and a lower bound for $e(M^n)$ where M^n is a Riemannian manifold of dimension n. (Guess: $n + 1$.)

If *no*, then the problem is

Problem 2′. Compute an upper bound for $e(M^n)$. (Guess: $n + 2$.)

The behavior of $e(M^n, r)$ for various r should give some *rough* idea of the "shape" of a Riemannian manifold M^n. For example, if $e(M^n, r) = n + 2$ we expect M^n to have a "spherical protrusion" of radius comparable to r/π as in Figure 4. Can you make this more precise?

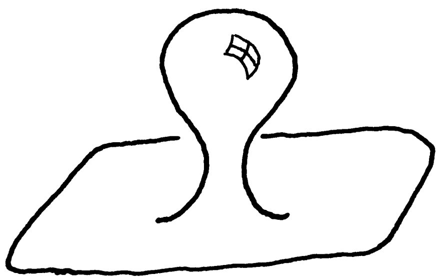


FIG. 4

Problem 3. Describe those Riemannian manifolds M^n for which

(a) $e(M^n) = n + 1, n + 2, \ldots,$ or

(b) $u(M^n) = 1, 2, \ldots, n + 1,$ or

(c) $e(M^n, r) = 1, 2, \ldots, n + 1, n + 2, \ldots$ for all r in D, some preassigned domain. In particular, what are the complete Riemannian manifolds M^n such that $e(M, r) = n + 1$ for all r?

Problem 4. Classify all complete metric spaces M with $e(M, r) = n + 1$ for all r. Or to what extent does $e(M, r) = n + 1$ characterize closed submetric spaces of the Riemannian examples found above?

Another good example of a problem where the problem is to decide what the problem is starts from R. P. Agnew, *Differential Equations*, McGraw-Hill, NY, 2nd ed., 1960, pp. 39–40:

> One day it started snowing at a heavy and steady rate. A snowplow started out at noon, going at 2 miles the first hour and 1 mile the second hour. What time did it start snowing?

To solve this, it is assumed that the plow clears a constant volume of snow per unit of time. This is unrealistic since the plow would move infinitely fast when there's no snow (though that doesn't arise in *this* problem). Perhaps there should also be an upper bound on the depth of snow the plow can cope with.

C. M. Bender, A. Toomre and Frederic Y. M. Wan [The snowplow problems, *Applied Math. Notes*, Vancouver, B.C. 1(1975) #1, 7–11] ask the additional question:

Suppose a second identical snowplow starts sometime after noon, say at $t = T$, in the wake of the first. If it ever catches up with the first plow, when, where and how?

The answer to "how" is "with infinite speed"! They also ask the question:

Three identical snowplows started at noon, 1 p.m. and 2 p.m., respectively. All three collided some time later. When did it start snowing?

The answer is given as 11:30 a.m. Answer all these questions under alternative assumptions. Also answer the following variant asked by M. G. Stone:

The snowplow starts out at noon. It turns around (instantaneously!) at 2 p.m. and returns along its own path at 3 p.m. What time did it start snowing? At what time should the plow have turned around to return by 2 p.m.?

IS A DISTANCE ONE PRESERVING MAPPING BETWEEN METRIC SPACES ALWAYS AN ISOMETRY?

THEMISTOCLES M. RASSIAS
4, Zagoras Street, Paradissos, Amaroussion, Athens, Greece

1. Preliminaries. We begin with the definition of an isometry: Let X, Y be two metric spaces, d_1, d_2 the distances on X and Y. A bijection mapping $f\colon X \to Y$, of X onto Y, is defined to be an **isometry** if $d_2(f(x), f(y)) = d_1(x, y)$ for all elements x, y of X. If $f\colon X \to Y$ is an isometry, then the inverse mapping $f^{-1}\colon Y \to X$ is an isometry of Y onto X. Two metric spaces X and Y are defined to be **isometric** if there exists an isometry of X onto Y. It thus follows that an isometry is an isomorphism for the metric space structures. Some of the properties of an isometry are mentioned in [**3**] and in [**4**]. We now state in which sense an incomplete space can be fattened out to be complete: If (X, d_1) is an incomplete metric space, then there exists a complete metric space $\tilde{X}$ so that X is isometric to a dense subset of $\tilde{X}$. Mazur and Ulam [**5**] have proved that every isometry of a normed real vector space onto a normed real vector space is a linear mapping up to translation. Consider then the following condition (distance one preserving property), for $f\colon X \to Y$.

(DOPP) *Given* $x, y \in X$ *with* $d_1(x, y) = 1$. *Then* $d_2(f(x), f(y)) = 1$.

2. The Problem. Let $f\colon X \to Y$ be a mapping (not necessarily continuous) satisfying condition (DOPP). Is $f\colon X \to f(X) \subseteq Y$ an isometry?

The problem still remains open even for the case where $X = R^n$ and $Y = R^m$ with $2 \leqslant n < m$ (see for example [**6**, p. 277]). Beadle [**1**] has covered a number of cases for mappings $f\colon R^n \to R^m$ that preserve some distance. For these mappings it is clear that $n \leqslant m$ because R^m has equilateral n-simplices if and only if $n \leqslant m$. Beckman and Quarles [**2**] proved that f is an isometry if $1 < n = m < \infty$. If $n < m$, f might not be an isometry if m is too large. It is not yet known if there is a unit distance preserving mapping $f\colon R^2 \to R^3$ which is not an isometry.

Acknowledgement. I am thankful to Prof. L. M. Kelly for helpful communication.

References

1. A. Beadle, Ph.D. Thesis, Michigan State University, 1977, p. 58, supervised by L. M. Kelly.
2. E. S. Beckman and D. A. Quarles, On isometries of Euclidean spaces, Proc. Amer. Math. Soc., 4 (1953) 810–815.
3. G. Choquet, Topology, Academic Press, New York and London, 1966.
4. J. Dieudonné, Foundations of Modern Analysis, Academic Press, New York and London, 1969.
5. S. Mazur et S. Ulam, Sur les Transformations Isométriques d'Espaces Vectoriels Normés, Comptes Rendus Acad. Sci., Paris, 194 (1932) 946–948.
6. J. Zaks, Problem No. 100, Contributions to Geometry, p. 277, Proceedings of the Geometry Symposium in Siegen, 1978, edited by J. Tölke and J. M. Wills, Birkhäuser Verlag Basel, 1979.

> Suppose a second identical snowplow starts sometime after noon, say at $t = T$, in the wake of the first. If it ever catches up with the first plow, when, where and how?

The answer to "how" is "with infinite speed"! They also ask the question:

> Three identical snowplows started at noon, 1 p.m. and 2 p.m., respectively. All three collided some time later. When did it start snowing?

The answer is given as 11:30 a.m. Answer all these questions under alternative assumptions. Also answer the following variant asked by M. G. Stone:

> The snowplow starts out at noon. It turns around (instantaneously!) at 2 p.m. and returns along its own path at 3 p.m. What time did it start snowing? At what time should the plow have turned around to return by 2 p.m.?

IS A DISTANCE ONE PRESERVING MAPPING BETWEEN METRIC SPACES ALWAYS AN ISOMETRY?

THEMISTOCLES M. RASSIAS
4, Zagoras Street, Paradissos, Amaroussion, Athens, Greece

1. Preliminaries. We begin with the definition of an isometry: Let X, Y be two metric spaces, d_1, d_2 the distances on X and Y. A bijection mapping f: $X \to Y$, of X onto Y, is defined to be an **isometry** if $d_2(f(x), f(y)) = d_1(x, y)$ for all elements x, y of X. If f: $X \to Y$ is an isometry, then the inverse mapping f^{-1}: $Y \to X$ is an isometry of Y onto X. Two metric spaces X and Y are defined to be **isometric** if there exists an isometry of X onto Y. It thus follows that an isometry is an isomorphism for the metric space structures. Some of the properties of an isometry are mentioned in [**3**] and in [**4**]. We now state in which sense an incomplete space can be fattened out to be complete: If (X, d_1) is an incomplete metric space, then there exists a complete metric space $\tilde{X}$ so that X is isometric to a dense subset of $\tilde{X}$. Mazur and Ulam [**5**] have proved that every isometry of a normed real vector space onto a normed real vector space is a linear mapping up to translation. Consider then the following condition (distance one preserving property), for f: $X \to Y$.

(DOPP) *Given* $x, y \in X$ *with* $d_1(x, y) = 1$. *Then* $d_2(f(x), f(y)) = 1$.

2. The Problem. Let f: $X \to Y$ be a mapping (not necessarily continuous) satisfying condition (DOPP). Is f: $X \to f(X) \subseteq Y$ an isometry?

The problem still remains open even for the case where $X = R^n$ and $Y = R^m$ with $2 \leqslant n < m$ (see for example [**6**, p. 277]). Beadle [**1**] has covered a number of cases for mappings f: $R^n \to R^m$ that preserve some distance. For these mappings it is clear that $n \leqslant m$ because R^m has equilateral n-simplices if and only if $n \leqslant m$. Beckman and Quarles [**2**] proved that f is an isometry if $1 < n = m < \infty$. If $n < m$, f might not be an isometry if m is too large. It is not yet known if there is a unit distance preserving mapping f: $R^2 \to R^3$ which is not an isometry.

Acknowledgement. I am thankful to Prof. L. M. Kelly for helpful communication.

References

1. A. Beadle, Ph.D. Thesis, Michigan State University, 1977, p. 58, supervised by L. M. Kelly.
2. E. S. Beckman and D. A. Quarles, On isometries of Euclidean spaces, Proc. Amer. Math. Soc., 4 (1953) 810–815.
3. G. Choquet, Topology, Academic Press, New York and London, 1966.
4. J. Dieudonné, Foundations of Modern Analysis, Academic Press, New York and London, 1969.
5. S. Mazur et S. Ulam, Sur les Transformations Isométriques d'Espaces Vectoriels Normés, Comptes Rendus Acad. Sci., Paris, 194 (1932) 946–948.
6. J. Zaks, Problem No. 100, Contributions to Geometry, p. 277, Proceedings of the Geometry Symposium in Siegen, 1978, edited by J. Tölke and J. M. Wills, Birkhäuser Verlag Basel, 1979.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

AN ELEMENTARY PROOF OF THE ISOMORPHISM $\mathbb{C}^* \approx S^1$

L. RICHARD DUFFY
33 Fellsway East, Malden, MA 02148

In his introductory algebra textbook [**1**], Louis Shapiro makes the following remark at the end of the section on group isomorphisms:

> Thus isomorphism is a simple and important concept. It is also complicated. For instance, only in the last few years were the nonzero complex numbers shown to be isomorphic to the complex numbers of absolute value 1.

This fact, which may seem a bit surprising (at least to beginning students of algebra), appears to have had its first published proof in [**2**].

Each of these two groups is a "divisible" group, meaning that the equation $x^n = g$ has a solution x in the group for every positive integer n and every g in the group. Using this fact, the proof cited above shows that the isomorphism in question is a simple corollary of the more general Structure Theorem for Divisible Abelian Groups (see [**3**], for example). While not excessively difficult, this latter theorem does take a while to establish from first principles, and its full power is by no means necessary to demonstrate the isomorphism claimed above. The following proof requires only the rudiments of group theory, together with Zorn's Lemma.

Let $\mathbb{C}^*$ denote the multiplicative group of nonzero complex numbers, S^1 its unit circle subgroup, $\mathbb{R}$ the additive group of reals, and $\mathbb{Z}$ the integers. The map sending $re^{i\theta}$ to the pair $(\log r, \theta/2\pi + \mathbb{Z})$ immediately establishes the isomorphisms $\mathbb{C}^* \approx \mathbb{R} \times \mathbb{R}/\mathbb{Z}$ and $S^1 \approx \mathbb{R}/\mathbb{Z}$. Thus it will be sufficient to show that $\mathbb{R} \times \mathbb{R}/\mathbb{Z} \approx \mathbb{R}/\mathbb{Z}$. To do this, we'll need three basic facts:

I. In a vector space, any independent set of elements can be enlarged to a basis. (Easily proved by means of Zorn's Lemma.)

II. If $\phi\colon G \to H$ is an isomorphism of groups and $A \lhd G$, then $\phi(A) \lhd H$ and the two quotient groups G/A and $H/\phi(A)$ are isomorphic.

III. If G, H are two groups with normal subgroups A, B, respectively, then $A \times B \lhd G \times H$, and $(G \times H)/(A \times B) \approx G/A \times H/B$.

To begin with, $\mathbb{R}$ is a vector space over the rational field $\mathbb{Q}$, and the singleton set $\{1\}$ is obviously independent over $\mathbb{Q}$. So by I, there is a basis $\mathfrak{B}$ of $\mathbb{R}$ over $\mathbb{Q}$ with $1 \in \mathfrak{B}$. Then the set $\mathfrak{B}^* = \mathfrak{B} \times \{0\} \cup \{0\} \times \mathfrak{B}$ is a basis of the vector space $\mathbb{R} \times \mathbb{R}$ over $\mathbb{Q}$ (in which the scalar multiplication is the obvious one given by $q \cdot (r, s) = (qr, qs)$). Furthermore, $\mathfrak{B}^*$ is a union of two disjoint sets, each equinumerous to the infinite set $\mathfrak{B}$, whence $\mathfrak{B}^*$ is equinumerous to $\mathfrak{B}$ (by a standard fact in set theory which is also proved by Zorn's Lemma).

Now let $\phi\colon \mathfrak{B}^* \to \mathfrak{B}$ be any one-one correspondence which maps the ordered pair $(0, 1)$ to 1. Then it is a simple matter to show that ϕ extends uniquely to a rational-vector-space isomorphism $\Phi\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, which is also a group isomorphism, of course. If n is any integer, then

$$\Phi((0,n)) = \Phi(n \cdot (0,1)) = n \cdot \Phi((0,1)) = n \cdot \phi((0,1)) = n \cdot 1 = n.$$

Thus Φ restricts to an isomorphism of the subgroup $\{0\} \times \mathbb{Z}$ onto $\mathbb{Z}$.

Now we're finished, because II gives $(\mathbb{R} \times \mathbb{R})/(\{0\} \times \mathbb{Z}) \approx \mathbb{R}/\mathbb{Z}$, and III gives $(\mathbb{R} \times \mathbb{R})/(\{0\} \times \mathbb{Z}) \approx \mathbb{R}/\{0\} \times \mathbb{R}/\mathbb{Z} \approx \mathbb{R} \times \mathbb{R}/\mathbb{Z}$, whence $\mathbb{R} \times \mathbb{R}/\mathbb{Z} \approx \mathbb{R}/\mathbb{Z}$, as desired.

References

1. L. Shapiro, Introduction to Abstract Algebra, McGraw-Hill, New York, 1975, p. 51.

2. J. R. Clay, The punctured plane is isomorphic to the unit circle, J. Number Theory, 1 (1969) 500–501.
3. J. J. Rotman, The Theory of Groups: an Introduction, 2nd ed., Allyn and Bacon, Boston, 1973, p. 186.

MATRICES WITH INTEGER ENTRIES AND INTEGER EIGENVALUES

J.- C. RENAUD
Mathematics Department, University of Papua New Guinea
Waigani, Papua New Guinea

When eigenvalues are first covered in linear algebra, it is often convenient to give numerically simple exercises—thus it may be desirable to provide problems which have integer solutions. It is reasonably easy to find examples of 2×2 and 3×3 matrices with integer entries and integer eigenvalues, but this leads naturally to the question: exactly how can one construct, in general, an $n \times n$ matrix with such a property? The theorem below gives a complete answer, in the sense that all such matrices can be built up in the same elementary way. Remarkably, its proof is based on only three well-known results.

THEOREM. *The $n \times n$ matrix A with integer entries has integer eigenvalues if and only if it is expressible in the form*

$$A = \sum_{i=1}^{n-1} \mathbf{u}_i^T \mathbf{v}_i + kI_n$$

where the $\mathbf{u}_i$ and $\mathbf{v}_i$ are row vectors with n integer components such that $\mathbf{u}_i \circ \mathbf{v}_j = 0$ for $1 \leqslant i < j \leqslant n-1$, k is an integer, and I_n is the unit $n \times n$ matrix.

The eigenvalues are $k, \mathbf{u}_1 \circ \mathbf{v}_1 + k, \dots, \mathbf{u}_{n-1} \circ \mathbf{v}_{n-1} + k$.

Discussion and Proof. Consider the $n \times n$ matrix A with integer entries and eigenvalues. We can choose one eigenvalue, say k, and form $B = A - kI_n$. Clearly B also has integer entries and eigenvalues, one of which is 0. It is now sufficient to show that B is expressible in the form of the summation term in A above. The first step is to show:

LEMMA 1. *Let B be an $n \times n$ matrix with integer entries whose determinant is zero. Then B is expressible in the form $B = XY$ where X is $n \times (n-1)$, Y is $(n-1) \times n$, and all entries are integers.*

Proof. This is based on a well-known result (see for example Theorem 7.10 in [**2**]): since the determinant of B is zero, B is expressible in the form $B = M_1 D M_2$ where M_1 and M_2 are invertible and D has the form diagonal $(d_1, \dots, d_{n-1}, 0)$, all entries of M_1, M_2, and D being integers. Now $D = D_1 D$ where $D_1 =$ diagonal $(1, \dots, 1, 0)$ is $n \times n$, and $B = (M_1 D_1)(D M_2)$. But the nth column of $M_1 D_1$ and the nth row of $D M_2$ are null: delete these, and call the remainder X and Y, respectively. Then $B = XY$ as required.

Note that the converse also holds.

Two other fundamental lemmas are required: Lemma 2 may be found in the exercises of Chapter 22 in [**1**], attributed to Barton, while Lemma 3 is Theorem III.12 in [**3**]. These are:

LEMMA 2. *Let $B = XY$ where X is $n \times (n-1)$, Y is $(n-1) \times n$. Then the characteristic polynomial of B, $|B - \lambda I_n|$, is equal to the polynomial $-\lambda|YX - \lambda I_{n-1}|$.*

LEMMA 3. *Let C be an $m \times m$ matrix with integer entries and integer eigenvalues. Then there exists T, $m \times m$ and invertible, $|T| = \pm 1$, with integer entries such that $T^{-1}CT$ is upper triangular.*

These three lemmas are sufficient to complete the proof of the theorem. We have $B = XY$, and there exists T, $(n-1) \times (n-1)$, such that $TYXT^{-1}$ is upper triangular. Moreover, the eigenvalues of XY are just 0 and those of YX, or those of $T(YX)T^{-1}$.

2. J. R. Clay, The punctured plane is isomorphic to the unit circle, J. Number Theory, 1 (1969) 500–501.
3. J. J. Rotman, The Theory of Groups: an Introduction, 2nd ed., Allyn and Bacon, Boston, 1973, p. 186.

MATRICES WITH INTEGER ENTRIES AND INTEGER EIGENVALUES

J.- C. RENAUD
Mathematics Department, University of Papua New Guinea
Waigani, Papua New Guinea

When eigenvalues are first covered in linear algebra, it is often convenient to give numerically simple exercises—thus it may be desirable to provide problems which have integer solutions. It is reasonably easy to find examples of 2×2 and 3×3 matrices with integer entries and integer eigenvalues, but this leads naturally to the question: exactly how can one construct, in general, an $n \times n$ matrix with such a property? The theorem below gives a complete answer, in the sense that all such matrices can be built up in the same elementary way. Remarkably, its proof is based on only three well-known results.

THEOREM. *The $n \times n$ matrix A with integer entries has integer eigenvalues if and only if it is expressible in the form*

$$A = \sum_{i=1}^{n-1} \mathbf{u}_i^T \mathbf{v}_i + kI_n$$

where the $\mathbf{u}_i$ and $\mathbf{v}_i$ are row vectors with n integer components such that $\mathbf{u}_i \circ \mathbf{v}_j = 0$ for $1 \leqslant i < j \leqslant n-1$, k is an integer, and I_n is the unit $n \times n$ matrix.

The eigenvalues are $k, \mathbf{u}_1 \circ \mathbf{v}_1 + k, \ldots, \mathbf{u}_{n-1} \circ \mathbf{v}_{n-1} + k$.

Discussion and Proof. Consider the $n \times n$ matrix A with integer entries and eigenvalues. We can choose one eigenvalue, say k, and form $B = A - kI_n$. Clearly B also has integer entries and eigenvalues, one of which is 0. It is now sufficient to show that B is expressible in the form of the summation term in A above. The first step is to show:

LEMMA 1. *Let B be an $n \times n$ matrix with integer entries whose determinant is zero. Then B is expressible in the form $B = XY$ where X is $n \times (n-1)$, Y is $(n-1) \times n$, and all entries are integers.*

Proof. This is based on a well-known result (see for example Theorem 7.10 in [**2**]): since the determinant of B is zero, B is expressible in the form $B = M_1 D M_2$ where M_1 and M_2 are invertible and D has the form diagonal $(d_1, \ldots, d_{n-1}, 0)$, all entries of M_1, M_2, and D being integers. Now $D = D_1 D$ where $D_1 =$ diagonal $(1, \ldots, 1, 0)$ is $n \times n$, and $B = (M_1 D_1)(D M_2)$. But the nth column of $M_1 D_1$ and the nth row of DM_2 are null: delete these, and call the remainder X and Y, respectively. Then $B = XY$ as required.

Note that the converse also holds.

Two other fundamental lemmas are required: Lemma 2 may be found in the exercises of Chapter 22 in [**1**], attributed to Barton, while Lemma 3 is Theorem III.12 in [**3**]. These are:

LEMMA 2. *Let $B = XY$ where X is $n \times (n-1)$, Y is $(n-1) \times n$. Then the characteristic polynomial of B, $|B - \lambda I_n|$, is equal to the polynomial $-\lambda|YX - \lambda I_{n-1}|$.*

LEMMA 3. *Let C be an $m \times m$ matrix with integer entries and integer eigenvalues. Then there exists T, $m \times m$ and invertible, $|T| = \pm 1$, with integer entries such that $T^{-1}CT$ is upper triangular.*

These three lemmas are sufficient to complete the proof of the theorem. We have $B = XY$, and there exists T, $(n-1) \times (n-1)$, such that $TYXT^{-1}$ is upper triangular. Moreover, the eigenvalues of XY are just 0 and those of YX, or those of $T(YX)T^{-1}$.

Now rewrite $B = (XT^{-1})(TY)$ and set $U = XT^{-1}$, $V = TY$. That is, $B = UV$ and VU is upper triangular. Letting $\mathbf{u}_i^T$ be the ith column of U and $\mathbf{v}_i$ be the ith row of V gives

$$B = \sum_{i=1}^{n-1} \mathbf{u}_i^T \mathbf{v}_i, \quad \text{and} \quad \mathbf{u}_i \circ \mathbf{v}_j = 0 \quad \text{for} \quad i < j.$$

Since the eigenvalues of B are those of VU, together with 0, these are 0 and $\mathbf{u}_i \circ \mathbf{v}_i$, $i = 1, \ldots, n-1$. Hence A is of the requisite form and has the stated property.

This is really only a proof of the "only if" case; however, the "if" case is merely a direct retracement of the above steps.

We give an example of this construction for a 4×4 case:

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} [1 \quad -2 \quad 1 \quad 1] + \begin{bmatrix} 2 \\ 1 \\ 2 \\ -2 \end{bmatrix} [-1 \quad -1 \quad 1 \quad 1] + \begin{bmatrix} -3 \\ 2 \\ 1 \\ 2 \end{bmatrix} [1 \quad 2 \quad -1 \quad 1]$$

$$+1 \cdot \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{bmatrix} = \begin{bmatrix} -3 & -10 & 6 & 0 \\ 1 & 4 & 3 & 3 \\ 0 & -2 & 3 & 4 \\ 4 & 6 & 0 & 1 \end{bmatrix}$$

has eigenvalues $1, 1 + 2, 1 + (-3), 1 + 2$ or $1, 3, -2, 3$.

REMARK. The author has used this technique in constructing exercises for students, and found it quite useful for 3×3 and 4×4 cases. However, the restriction $\mathbf{u}_i \circ \mathbf{v}_j = 0$ for $i < j$ makes it clumsy at higher levels (which are in any case beyond classroom work).

Problem. It would be pleasant to be able also to give a nice formula for the eigenvectors. For the 3×3 case, we obtain

eigenvalue	eigenvector
k	$\mathbf{v}_1 \times \mathbf{v}_2$
$k + \mathbf{u}_1 \circ \mathbf{v}_1$	$\mathbf{u}_1$
$k + \mathbf{u}_2 \circ \mathbf{v}_2$	$\mathbf{v}_1 \times (\mathbf{u}_1 \times \mathbf{u}_2) + (\mathbf{u}_2 \circ \mathbf{v}_2)\mathbf{u}_2$
	$= (\mathbf{u}_2 \circ \mathbf{v}_2 - \mathbf{u}_1 \circ \mathbf{v}_1)\mathbf{u}_2 + (\mathbf{u}_2 \circ \mathbf{v}_1)\mathbf{u}_1$

This is almost neat, but probably will not generalize easily.

References

1. J. W. Archbold, Algebra, Pitman, London, 1964.
2. B. Hartley and T. O. Hawkes, Rings, Modules and Linear Algebra, Chapman and Hall, London, 1970.
3. M. Newman, Integral Matrices, Academic Press, New York, 1972.

WHEN IS $L^p(\mu)$ CONTAINED IN $L^q(\mu)$?

JUAN L. ROMERO

Departamento de Matemáticas, Universidad de Sevilla, C / Tarfia, s / n, Sevilla, Spain

Throughout this note, Ω will denote a nonempty set, Σ a σ-algebra of subsets of Ω, and μ a positive measure on the measurable space (Ω, Σ). We derive characterizations of the measure spaces (Ω, Σ, μ) such that $0 < p < q$ implies either $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$, thus completing and improving some results of B. Subramanian [**1**].

Now rewrite $B = (XT^{-1})(TY)$ and set $U = XT^{-1}$, $V = TY$. That is, $B = UV$ and VU is upper triangular. Letting $\mathbf{u}_i^T$ be the ith column of U and $\mathbf{v}_i$ be the ith row of V gives

$$B = \sum_{i=1}^{n-1} \mathbf{u}_i^T \mathbf{v}_i, \quad \text{and} \quad \mathbf{u}_i \circ \mathbf{v}_j = 0 \quad \text{for} \quad i < j.$$

Since the eigenvalues of B are those of VU, together with 0, these are 0 and $\mathbf{u}_i \circ \mathbf{v}_i$, $i = 1, \ldots, n-1$. Hence A is of the requisite form and has the stated property.

This is really only a proof of the "only if" case; however, the "if" case is merely a direct retracement of the above steps.

We give an example of this construction for a 4×4 case:

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} [1 \quad -2 \quad 1 \quad 1] + \begin{bmatrix} 2 \\ 1 \\ 2 \\ -2 \end{bmatrix} [-1 \quad -1 \quad 1 \quad 1] + \begin{bmatrix} -3 \\ 2 \\ 1 \\ 2 \end{bmatrix} [1 \quad 2 \quad -1 \quad 1]$$

$$+1 \cdot \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{bmatrix} = \begin{bmatrix} -3 & -10 & 6 & 0 \\ 1 & 4 & 3 & 3 \\ 0 & -2 & 3 & 4 \\ 4 & 6 & 0 & 1 \end{bmatrix}$$

has eigenvalues 1, $1 + 2$, $1 + (-3)$, $1 + 2$ or $1, 3, -2, 3$.

REMARK. The author has used this technique in constructing exercises for students, and found it quite useful for 3×3 and 4×4 cases. However, the restriction $\mathbf{u}_i \circ \mathbf{v}_j = 0$ for $i < j$ makes it clumsy at higher levels (which are in any case beyond classroom work).

Problem. It would be pleasant to be able also to give a nice formula for the eigenvectors. For the 3×3 case, we obtain

eigenvalue	eigenvector
k	$\mathbf{v}_1 \times \mathbf{v}_2$
$k + \mathbf{u}_1 \circ \mathbf{v}_1$	$\mathbf{u}_1$
$k + \mathbf{u}_2 \circ \mathbf{v}_2$	$\mathbf{v}_1 \times (\mathbf{u}_1 \times \mathbf{u}_2) + (\mathbf{u}_2 \circ \mathbf{v}_2)\mathbf{u}_2$
	$= (\mathbf{u}_2 \circ \mathbf{v}_2 - \mathbf{u}_1 \circ \mathbf{v}_1)\mathbf{u}_2 + (\mathbf{u}_2 \circ \mathbf{v}_1)\mathbf{u}_1$

This is almost neat, but probably will not generalize easily.

References

1. J. W. Archbold, Algebra, Pitman, London, 1964.
2. B. Hartley and T. O. Hawkes, Rings, Modules and Linear Algebra, Chapman and Hall, London, 1970.
3. M. Newman, Integral Matrices, Academic Press, New York, 1972.

WHEN IS $L^p(\mu)$ CONTAINED IN $L^q(\mu)$?

JUAN L. ROMERO

Departamento de Matemáticas, Universidad de Sevilla, C / Tarfia, s / n, Sevilla, Spain

Throughout this note, Ω will denote a nonempty set, Σ a σ-algebra of subsets of Ω, and μ a positive measure on the measurable space (Ω, Σ). We derive characterizations of the measure spaces (Ω, Σ, μ) such that $0 < p < q$ implies either $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$, thus completing and improving some results of B. Subramanian [**1**].

DEFINITION. A μ-measurable set $E \subset \Omega$ such that $0 < \mu(E) \leqslant +\infty$ is called an atom whenever for any μ-measurable subset E_1 of E we have either $\mu(E_1) = 0$ or $\mu(E_1) = \mu(E)$.

PROPOSITION 0.I [**2**, pg. 67]. *If E is a set of positive measure such that neither E nor any of its subsets is an atom, and $0 < \alpha < \mu(E)$, then there exists a subset E_1 of E such that $\mu(E_1) = \alpha$.*

PROPOSITION 0.II [**2**, pg. 67]. *If E is a set of σ-finite μ-measure, then $E = E_1 \cup E_2$ uniquely, where neither E_1 nor any of its measurable subsets is an atom and E_2 is a union of an at most countable number of atoms of finite measure.*

With these preliminaries, we have:

THEOREM 1. *The following conditions on the measure space (Ω, Σ, μ) are equivalent:*

(i) *$L^p(\mu) \subset L^q(\mu)$ for some pair p, q of positive numbers such that $p < q$.*

(ii) *There does not exist a disjoint sequence $\{E_n\}$ of sets of positive measure in Ω such that the sequence $\{\mu(E_n)\}$ tends to zero.*

(iii) *Every set of finite measure in Ω is a finite pairwise disjoint union of atoms, and the measures of the atoms in Ω are bounded away from zero.*

(iv) *$L^p(\mu) \subset L^q(\mu)$ for all positive real numbers p and q such that $p < q$.*

Proof. (i) $\Rightarrow$ (ii). Suppose that there exists a disjoint sequence $\{E_n\}$ of sets of positive measure in Ω such that the sequence $\{\mu(E_n)\}$ tends to zero. It can be assumed that the sequence $\{\mu(E_n)\}$ is strictly decreasing and that $\mu(E_n) < 1$ for every $n \in \mathbb{N}$. Denote by I_k the interval $[k^{-\alpha}, (k-1)^{-\alpha})$, where $\alpha = (3/2)q/(q-p)$, $k = 2, 3, \ldots$. One can construct, inductively, two strictly increasing sequences of positive integers $\{h_n\}$ and $\{k_n\}$ such that $\mu(E_{h_n}) \in I_{k_n}$ for every $n \in \mathbb{N}$. Let h: $\Omega \to \mathbb{R}$ be the function defined as follows:

$$h(x) = k_n^{\beta} \text{ when } x \in E_{h_n}$$

$$h(x) = 0 \text{ when } x \in \Omega \setminus \bigcup_{n=1}^{\infty} E_{h_n},$$

where $\beta = (3/2)/(q-p)$.

We have:

$$\int_{\Omega} h^p d\mu = \sum_{n=1}^{\infty} k_n^{\beta p} \mu(E_{h_n}) < \sum_{n=1}^{\infty} k_n^{\beta p}(k_n - 1)^{-\alpha} < +\infty$$

$$\int_{\Omega} h^q d\mu = \sum_{n=1}^{\infty} k_n^{\beta q} \mu(E_{h_n}) \geqslant \sum_{n=1}^{\infty} k_n^{\beta q} k_n^{-\alpha} = +\infty.$$

This contradiction proves (ii).

(ii) $\Rightarrow$ (iii). According to Proposition 0.II, if $E \subset \Omega$ is a set of finite measure, one can write $E = E_1 \cup E_2$, where E_1 is atom free and E_2 is purely atomic. If $\mu(E_1) > 0$, we can construct, by Proposition 0.I, a sequence $\{F_n\}$ of measurable sets such that $F_1 = E_1 \supset F_2 \supset \cdots \supset F_n \supset \cdots$ and $\mu(F_k) = \mu(E_1)/2^{k-1}$ $(k = 1, 2, \ldots)$. For $n \in \mathbb{N}$, define $G_n = F_n \setminus F_{n+1}$. The sequence $\{G_n\}$ is a disjoint sequence of sets of positive measure in Ω such that $\mu(G_n) \to 0$ as $n \to \infty$. This contradicts (ii). Hence, one can write $E = \{\cup_{i \in I} A_i\} \cup E_0$, where $A_i \in \Sigma$ is, for every $i \in I$, an atom and E_0 is such that $\mu(E_0) = 0$. Since $\mu(E) < +\infty$, the set I is, at most, countable. If the collection of positive real numbers $\{\mu(A_i), i \in I\}$ is not bounded away from zero, it would contradict (ii). Hence I is a finite set. If there is no number $m > 0$ such that the set of values of μ is contained in $\{0\} \cup [m, \infty]$, we could find atoms in Ω whose measure is arbitrarily small, which contradicts (ii). This proves (iii).

(iii) $\Rightarrow$ (iv). Assume f: $\Omega \to \mathbb{R}$ is a function such that $a = \int_{\Omega} |f|^p \, d\mu < +\infty$. For every $n \in \mathbb{N}$, we set $E_n = \{x \in \Omega : |f(x)| > n\}$. Now

$$n^p \mu(E_n) \leqslant \int_{E_n} |f|^p \, d\mu \leqslant \int_{\Omega} |f|^p \, d\mu = a < +\infty,$$

so $\mu(E_n) < +\infty$ for every $n \in \mathbb{N}$, and $\mu(E_n) \to 0$ as $n \to \infty$. Condition (iii) implies that there exists an $n_0 \in \mathbb{N}$ such that $\mu(E_n) = 0$ for $n \geqslant n_0$. This proves that $|f(x)| \leqslant n_0$ for all $x \in \Omega$ except a set of μ-measure zero. Hence

$$\int_\Omega |f|^q \, d\mu = \int_\Omega |f|^p \cdot |f|^{q-p} \, d\mu \leqslant n_0^{q-p} \int_\Omega |f|^p \, d\mu < +\infty.$$

This proves (iv).

The proof of (iv) $\Rightarrow$ (i) is obvious. Q.E.D.

THEOREM 2. *The following conditions on the measure space* (Ω, Σ, μ) *are equivalent*:

(I) $L^q(\mu) \subset L^p(\mu)$ *for some pair* p, q *of positive numbers such that* $p < q$.

(II) *Every disjoint sequence* $\{E_n\}$ *of sets of positive finite measure in* Ω *is such that the sequence* $\{\mu(E_n)\}$ *is bounded.*

(III) $\Omega = E_1 \cup E_2$, *where* $E_1 \subset \Omega$ *is a measurable set such that* $\mu(E_1) < +\infty$ *and* E_2 *is either empty or an atom of infinite measure.*

(IV) $L^q(\mu) \subset L^p(\mu)$ *for all positive numbers p and q such that* $p < q$.

Proof. (I) $\Rightarrow$ (II). Suppose that there exists a disjoint sequence $\{E_n\}$ of sets of positive measure such that the sequence $\{\mu(E_n)\}$ diverges to $+\infty$. One can assume that the sequence $\{\mu(E_n)\}$ is strictly increasing and that $\mu(E_n) > 1$ for every $n \in \mathbb{N}$. Denote by I_k the interval $(2^k, 2^{k+1}]$, for every $k \in \mathbb{N}$. One can construct, inductively, two strictly increasing sequences of positive integers $\{h_n\}$ and $\{k_n\}$ such that $\mu(E_{h_n}) \in I_{k_n}$, for every $n \in \mathbb{N}$. Let $g\colon \Omega \to \mathbb{R}$ be the function defined as follows:

$$g(x) = 2^{-k_n/p} \text{ when } x \in E_{h_n},\ g(x) = 0 \text{ when } x \in \Omega \setminus \bigcup_{n=1}^{\infty} E_{h_n}.$$

We have that $\int |g|^q \, d\mu < +\infty$ and $\int |g|^p \, d\mu = +\infty$. This contradiction proves (II).

(II) $\Rightarrow$ (III). Let $\beta = \sup \mu(F)$, where the sup is taken over all measurable subsets F of Ω satisfying $\mu(F) < \infty$. There exists a sequence $\{F_n\}$ of measurable subsets of Ω such that $\lim_{n\to\infty} \mu(F_n) = \beta$. Let $E_1 = \cup_{n=1}^{\infty} F_n$, so $\mu(E_1) = \beta$. Condition (II) implies that $\beta < \infty$. Let $E_2 = \Omega \setminus E_1$. If $E \subset E_2$ is such that $\mu(E) < \infty$, the construction of E_1 proves that $\mu(E) = 0$. This shows that E_2 is either empty or an infinite atom and (III) is proved.

(III) $\Rightarrow$ (IV). Assume $f\colon \Omega \to \mathbb{R}$ is a measurable function such that $\int_\Omega |f|^q \, d\mu < +\infty$. If we set

$$A_n = \left\{ x \in E_2 \colon |f(x)| > \frac{1}{n} \right\},$$

we have either $\mu(A_n) = 0$ or $\mu(A_n) = +\infty$. Since

$$\frac{1}{n^q} \mu(A_n) \leqslant \int_{A_n} |f|^q \, d\mu \leqslant \int_\Omega |f|^q \, d\mu < +\infty,$$

it follows that $\mu(A_n) = 0$ for every $n \in \mathbb{N}$. Let A be the set $\cup_{n=1}^{\infty} A_n = \{x \in E_2 \colon f(x) \neq 0\}$. We have $\mu(A) = 0$ and

$$\int_\Omega |f|^q \, d\mu = \int_{E_1} |f|^q \, d\mu < +\infty.$$

Since $\mu(E_1) < +\infty$, Hölder's inequality implies that

$$\int_\Omega |f|^p \, d\mu = \int_{E_1} |f|^p \, d\mu \leqslant \left(\int_{E_1} (|f|^p)^{q/p} \right)^{p/q} \left(\int_{E_1} 1 \, d\mu \right)^{1-p/q}$$

$$= \mu(E_1)^{1-p/q} \left(\int_{E_1} |f|^q \, d\mu \right)^{p/q} < +\infty.$$

This proves (IV). The proof (IV) $\Rightarrow$ (I) is obvious. Q.E.D.

References

1. B. Subramanian, On the inclusion $L^p(\mu) \subset L^q(\mu)$, this MONTHLY, 85 (1978) 479–481.
2. A. C. Zaanen, Integration, North-Holland, Amsterdam, 1967.

A SHORT PROOF OF THE VARIATIONAL PRINCIPLE FOR APPROXIMATE SOLUTIONS OF A MINIMIZATION PROBLEM

J.- B. HIRIART - URRUTY
U.E.R. de Mathématiques, Université Paul Sabatier, 118, route de Narbonne, 31062 TOULOUSE Cédex, France

Given any problem of minimization, two tasks come up very naturally: the *question* of the existence of a minimum point and the determination of *properties* (usually in terms of conditions for optimality) satisfied by minimum points. The classical variational principle states that if a differentiable function attains its minimum at some point x_0, then $f'(x_0) = 0$. When f has only a finite lower bound (which it does not necessarily attain), then, for every $\varepsilon > 0$, there exists some *approximate solution*, i.e., a point x_ε such that:

$$\inf f \leqslant f(x_\varepsilon) \leqslant \inf f + \varepsilon. \tag{1}$$

What conditions do such points necessarily satisfy? To answer this question is precisely the purpose of Ekeland's variational principle [**1**]. Since its introduction in 1972, this variational principle has received a great deal of attention: different aspects and applications of it as well as a general proof are reported in [**2**]. The aim of this note is to present a short proof which relies on the regularization method applied to the original problem [**3**]. Although the proof essentially works only in finite dimensions, it is appealing for its geometrical insight and the simplicity of techniques involved.

THEOREM. *Let f: $\mathbb{R}^n \to (-\infty, +\infty]$ be a lower semicontinuous function, not identically $+\infty$, bounded from below. Then for every point x_ε satisfying* (1) *and every $\lambda > 0$, there exists some point $\bar{x}_\varepsilon$ such that*

$$f(\bar{x}_\varepsilon) \leqslant f(x_\varepsilon), \tag{2}$$

$$\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant \lambda, \tag{3}$$

$$\forall x \in \mathbb{R}^n, f(\bar{x}_\varepsilon) \leqslant f(x) + (\varepsilon/\lambda)\|x - \bar{x}_\varepsilon\|. \tag{4}$$

Proof. Given a point x_ε satisfying (1), we consider the perturbed function

$$g\colon x \to f(x) + (\varepsilon/\lambda)\|x - x_\varepsilon\|.$$

Since f is assumed lower semicontinuous and bounded from below, g is clearly lower semicontinuous and verifies: $\lim_{\|x\| \to +\infty} g(x) = +\infty$. Therefore there exists $\bar{x}_\varepsilon$ minimizing g on $\mathbb{R}^n$ such that

$$\forall x \in \mathbb{R}^n, f(\bar{x}_\varepsilon) + (\varepsilon/\lambda)\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant f(x) + (\varepsilon/\lambda)\|x - x_\varepsilon\|. \tag{5}$$

By letting $x = x_\varepsilon$ we get

$$f(\bar{x}_\varepsilon) + (\varepsilon/\lambda)\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant f(x_\varepsilon)$$

and (2) follows. Now, since $f(x_\varepsilon) \leqslant \inf f + \varepsilon$, we clearly deduce from the above that $\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant \lambda$.

We infer from (5) that

$$\begin{aligned} f(\bar{x}_\varepsilon) &\leqslant f(x) + (\varepsilon/\lambda)\{\|x - x_\varepsilon\| - \|\bar{x}_\varepsilon - x_\varepsilon\|\} \\ &\leqslant f(x) + (\varepsilon/\lambda)\|x - \bar{x}_\varepsilon\| \quad \text{for all } x, \end{aligned}$$

which is the desired inequality (4). □

References

1. B. Subramanian, On the inclusion $L^p(\mu) \subset L^q(\mu)$, this MONTHLY, 85 (1978) 479–481.
2. A. C. Zaanen, Integration, North-Holland, Amsterdam, 1967.

A SHORT PROOF OF THE VARIATIONAL PRINCIPLE FOR APPROXIMATE SOLUTIONS OF A MINIMIZATION PROBLEM

J.- B. HIRIART - URRUTY
U.E.R. de Mathématiques, Université Paul Sabatier, 118, route de Narbonne, 31062 TOULOUSE Cédex, France

Given any problem of minimization, two tasks come up very naturally: the *question* of the existence of a minimum point and the determination of *properties* (usually in terms of conditions for optimality) satisfied by minimum points. The classical variational principle states that if a differentiable function attains its minimum at some point x_0, then $f'(x_0) = 0$. When f has only a finite lower bound (which it does not necessarily attain), then, for every $\varepsilon > 0$, there exists some *approximate solution*, i.e., a point x_ε such that:

(1) $$\inf f \leqslant f(x_\varepsilon) \leqslant \inf f + \varepsilon.$$

What conditions do such points necessarily satisfy? To answer this question is precisely the purpose of Ekeland's variational principle [**1**]. Since its introduction in 1972, this variational principle has received a great deal of attention: different aspects and applications of it as well as a general proof are reported in [**2**]. The aim of this note is to present a short proof which relies on the regularization method applied to the original problem [**3**]. Although the proof essentially works only in finite dimensions, it is appealing for its geometrical insight and the simplicity of techniques involved.

THEOREM. *Let f: $\mathbb{R}^n \to (-\infty, +\infty]$ be a lower semicontinuous function, not identically $+\infty$, bounded from below. Then for every point x_ε satisfying* (1) *and every $\lambda > 0$, there exists some point $\bar{x}_\varepsilon$ such that*

(2) $$f(\bar{x}_\varepsilon) \leqslant f(x_\varepsilon),$$

(3) $$\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant \lambda,$$

(4) $$\forall x \in \mathbb{R}^n, f(\bar{x}_\varepsilon) \leqslant f(x) + (\varepsilon/\lambda)\|x - \bar{x}_\varepsilon\|.$$

Proof. Given a point x_ε satisfying (1), we consider the perturbed function

$$g\colon x \to f(x) + (\varepsilon/\lambda)\|x - x_\varepsilon\|.$$

Since f is assumed lower semicontinuous and bounded from below, g is clearly lower semicontinuous and verifies: $\lim_{\|x\| \to +\infty} g(x) = +\infty$. Therefore there exists $\bar{x}_\varepsilon$ minimizing g on $\mathbb{R}^n$ such that

(5) $$\forall x \in \mathbb{R}^n, f(\bar{x}_\varepsilon) + (\varepsilon/\lambda)\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant f(x) + (\varepsilon/\lambda)\|x - x_\varepsilon\|.$$

By letting $x = x_\varepsilon$ we get

$$f(\bar{x}_\varepsilon) + (\varepsilon/\lambda)\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant f(x_\varepsilon)$$

and (2) follows. Now, since $f(x_\varepsilon) \leqslant \inf f + \varepsilon$, we clearly deduce from the above that $\|\bar{x}_\varepsilon - x_\varepsilon\| \leqslant \lambda$.

We infer from (5) that

$$\begin{aligned} f(\bar{x}_\varepsilon) &\leqslant f(x) + (\varepsilon/\lambda)\{\|x - x_\varepsilon\| - \|\bar{x}_\varepsilon - x_\varepsilon\|\} \\ &\leqslant f(x) + (\varepsilon/\lambda)\|x - \bar{x}_\varepsilon\| \quad \text{for all } x, \end{aligned}$$

which is the desired inequality (4). □

Thus, the closer to x_ε we desire $\bar{x}_\varepsilon$ to be, the larger the perturbation of f that must be accepted. A good compromise is to take $\lambda = \varepsilon^{1/2}$. As an illustration, we consider what the variational principle for approximate solutions yields for differentiable f.

COROLLARY. *Let f: $\mathbb{R}^n \to \mathbb{R}$ be a differentiable function, bounded from below. For every $\varepsilon > 0$, there exists some point $\bar{x}_\varepsilon$ satisfying*:

$$f(\bar{x}_\varepsilon) \leqslant \inf f + \varepsilon,$$
$$\|f'(\bar{x}_\varepsilon)\| \leqslant \varepsilon^{1/2}.$$

References

1. I. Ekeland, On the variational principle, J. Math. Anal. Appl., 47 (1974) 324-353.
2. ______, Nonconvex minimization problems, Bull. Amer. Math. Society, 1 (1979) 443–474.
3. P. Loridan, Solutions approchées de problèmes d'optimisation, Communication au Colloque d'Analyse Numérique, Imbours (1977).

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

THE TRUE GROWTH RATE AND THE INFLATION BALANCING PRINCIPLE

ROBERT C. THOMPSON
Mathematics Department, University of California, Santa Barbara, CA 93106

Twenty-five or so years ago, mathematics of finance was a staple in the undergraduate curriculum. However, it fell into disfavor, owing chiefly to the tedious nature of the required computations, and perhaps also owing to a lack of interest in it on the part of mathematics faculty. There was, however, no scientific reason for this fall from grace—finance is an important and tidy branch of applied mathematics. With a reawakening awareness of applied mathematics, and the change in the computational aspect from tedious to exciting created by the pocket calculator, financial mathematics may recover some importance in faculty eyes. But the traditional presentation needs modernizing.

Beyond any doubt, the most important change needed is to recognize that it no longer is adequate to think in terms of interest alone: inflation must be brought squarely into the picture. To this end, there are two theorems, small ones to be sure, but not so small to be immediately obvious, that greatly facilitate the mathematical treatment of inflation within the context of finance. The purpose of this note is to present these two theorems, together with a supplement explaining their connection with the mathematics of inflation as it usually is given in economics. The author has repeatedly used these theorems in the finance classes that he regularly teaches, with very satisfactory student comprehension.

The basic issue to explore in a treatment of inflation is the *true growth rate*. If i is the interest rate, and r the inflation rate, most people would assert that the true growth rate simply is

$$i - r = \text{interest rate} - \text{inflation rate}.$$

An economist, upon being asked about this formula, would probably say "Of course, it's correct!" A mathematician, upon being asked, might agree that it is correct, though possibly with an

Supplement. Having now thoroughly convinced the reader that the formula $g = i - r$ for the true growth rate is wrong, we now surprise him by asserting that it is correct (but only when properly formulated.)

THEOREM 3. *In an inflationary economy, with i the interest rate and r the inflation rate, both under continuous compounding, the true growth rate is*

$$g = i - r.$$

Proof. The reasoning is the same as in Theorem 1, except that, if the time span is T years (instead of n periods), the future values Q and Y are

$$Q = Pe^{iT}, \qquad Y = Xe^{rT},$$

under interest   under inflation.

Setting $Y = Q$ then produces $X = Pe^{(i-r)T}$, but we also have $X = Pe^{gT}$ under continuously compounded true growth. So $g = i - r$.

In finance, interest is usually presented on a periodically compounded basis. It is true that continuously compounded rates have sometimes occurred in commercial practice, but invariably the equivalent annually compounded figures have also been prominently mentioned. Inflation, though, seems never to be presented to the public on any basis other than monthly compounding or annual compounding. (Percents per month inflation, or percents per year, regularly appear in the popular press.) In financial work, therefore, the periodically compounded true growth rate formula $g = (i - r)/(1 + r)$ is the one to use. However, in economics, the presentation generally is more "theoretical," and the compounding is continuous. Thus, it is perfectly in order for the mathematician to tell the financier that the formula $g = i - r$ for the true growth rate is wrong, and yet tell the economist that it is right!

Our final remark is that financial mathematics is a beautiful subject when it is made exciting by incorporating inflation into the picture, provided that hand-held calculators (*not* compound interest tables) are used for the computations. The author has several years experience to demonstrate this, teaching the subject to large classes of students, all taking the course as an elective. Moreover, the mathematics graduate students who have been the author's assistants with the course all speak extremely favorably of it, even though almost all found that considerable effort was needed to master the material. For them, it is an important and not so trivial branch of applied mathematics. The newest feature, still not very much recognized, is that incorporating inflation makes it easy to do computations with startling and unsettling conclusions.

A TECHNIQUE FOR INTEGRATION BY PARTS

HERBERT E. KASUBE
Department of Mathematics, Bradley University, Peoria, IL 61625

When we first encounter integration by parts we are confronted with the usual formula:

$$\int u\, dv = uv - \int v\, du.$$

The expressions u and dv are chosen with two criteria in mind:

(1) v should be easy to find from dv;
(2) the integral $\int v\, du$ should, in some sense, be "better" or easier than $\int u\, dv$.

The difficulty usually arises in choosing u and dv to satisfy these conditions. An acronym gives us a method of selection which helps to satisfy criterion (2) above. The word is LIATE, standing for *L*ogarithmic *I*nverse trig *A*lgebraic *T*rig *E*xponential. These five basic types of functions are those encountered in our integral.

If, for example, our integral involves the product of a logarithmic function and an algebraic function, we choose u to be the log part and, keeping criterion (1) in mind, the algebraic part is chosen for dv. By making this selection, in $\int v\,du$, the logarithmic character has disappeared by virtue of the differentiation. The antidifferentiation of the algebraic part results (most of the time) in an algebraic v. Consequently, the integral $\int v\,du$ is strictly algebraic in its integrand and hopefully easier than our original integral.

Similarly, if we have any combination of two of these types of functions in our original integral, we choose for u that type that appears first in LIATE and dv is whatever is left, that part that appears second in LIATE.

Rationale: Differentiating logs and inverse trig functions changes their character to algebraic, while algebraic functions run the "middle of the road;" differentiating them yields algebraic, while integration sometimes yields nonalgebraic, so we would probably differentiate them before integrating. Trig and exponential functions "bring up the rear" since their differentiation or antidifferentiation results in similar functions.

As an example, consider $\int x\ln(x)\,dx$. Using LIATE, we pick $u = \ln(x)$ and $dv = x\,dx$. We obtain $du = x^{-1}\,dx$ and $v = x^2/2$. Consequently,

$$\begin{aligned}\int x\ln(x)\,dx &= \frac{x^2}{2}\ln(x) - \int \frac{x}{2}\,dx\\ &= \frac{x^2}{2}\ln(x) - \frac{x^2}{4} + C.\end{aligned}$$

Students have liked using LIATE and find that choosing u and dv is much less frightening. The author has yet to see an integral done by integration by parts to which LIATE cannot be successfully applied.

To be able to use integration by parts, perhaps it is better LIATE than never.

Reference

1. Howard Anton, Calculus with Analytic Geometry, Wiley, New York, 1980.

MISCELLANEA

95. For the most part [Glaisher] accepted or rejected contributions on his own responsibility. The best editors have of course a wonderful power of judging the quality of a manuscript without attempting to read it; but the task becomes steadily more difficult as mathematics grows and specializes... A private journal, controlled autocratically by a kindly and discriminating editor, can serve many useful purposes, and particularly that of giving early encouragement to beginners, on whom societies, with councils and referees, are sometimes very severe.

—G. H. Hardy, "Dr. Glaisher and the 'Messenger of Mathematics,'" *Messenger of Mathematics*, 58 (1929) 159–160.

ANSWER TO PHOTOS ON PAGE 195.

Top: P. S. Alexandroff and H. Hopf (topology). Bottom: A. D. Alexandroff and E. Hopf (analysis, of two different but overlapping kinds).

If, for example, our integral involves the product of a logarithmic function and an algebraic function, we choose u to be the log part and, keeping criterion (1) in mind, the algebraic part is chosen for dv. By making this selection, in $\int v\,du$, the logarithmic character has disappeared by virtue of the differentiation. The antidifferentiation of the algebraic part results (most of the time) in an algebraic v. Consequently, the integral $\int v\,du$ is strictly algebraic in its integrand and hopefully easier than our original integral.

Similarly, if we have any combination of two of these types of functions in our original integral, we choose for u that type that appears first in LIATE and dv is whatever is left, that part that appears second in LIATE.

Rationale: Differentiating logs and inverse trig functions changes their character to algebraic, while algebraic functions run the "middle of the road;" differentiating them yields algebraic, while integration sometimes yields nonalgebraic, so we would probably differentiate them before integrating. Trig and exponential functions "bring up the rear" since their differentiation or antidifferentiation results in similar functions.

As an example, consider $\int x\ln(x)\,dx$. Using LIATE, we pick $u = \ln(x)$ and $dv = x\,dx$. We obtain $du = x^{-1}\,dx$ and $v = x^2/2$. Consequently,

$$\begin{aligned}\int x\ln(x)\,dx &= \frac{x^2}{2}\ln(x) - \int \frac{x}{2}\,dx\\ &= \frac{x^2}{2}\ln(x) - \frac{x^2}{4} + C.\end{aligned}$$

Students have liked using LIATE and find that choosing u and dv is much less frightening. The author has yet to see an integral done by integration by parts to which LIATE cannot be successfully applied.

To be able to use integration by parts, perhaps it is better LIATE than never.

Reference

1. Howard Anton, Calculus with Analytic Geometry, Wiley, New York, 1980.

MISCELLANEA

95. For the most part [Glaisher] accepted or rejected contributions on his own responsibility. The best editors have of course a wonderful power of judging the quality of a manuscript without attempting to read it; but the task becomes steadily more difficult as mathematics grows and specializes... A private journal, controlled autocratically by a kindly and discriminating editor, can serve many useful purposes, and particularly that of giving early encouragement to beginners, on whom societies, with councils and referees, are sometimes very severe.

—G. H. Hardy, "Dr. Glaisher and the 'Messenger of Mathematics,'" *Messenger of Mathematics*, 58 (1929) 159–160.

ANSWER TO PHOTOS ON PAGE 195.

Top: P. S. Alexandroff and H. Hopf (topology). Bottom: A. D. Alexandroff and E. Hopf (analysis, of two different but overlapping kinds).

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2989. *Proposed by M. Goldberg, University of Waterloo, and S. C. Locke, Florida Atlantic University.*

Let $A = (a, a')$ and $B = (b, b')$ be lattice points (points in the plane with integral coordinates), and let $d(A, B)$ denote $|a - b| + |a' - b'|$. Let S be the set of points at d-distance at most k from the origin. Calculate

$$f(k) = \sum_{A,B\in S} d(A, B).$$

E 2990. *Proposed by H. Eves, University of Maine, and C. Kimberling, University of Evansville.*

Let ABC be a triangle and L a line in the plane of ABC not passing through A, B, or C.

i) Prove that the isogonal conjugate of L is an ellipse, parabola or hyperbola according as L meets the circumcircle of ABC in zero, one or two points.
ii) Prove that the isotomic conjugate of L is an ellipse, parabola, or hyperbola according as L meets E in zero, one, or two points, where E is the ellipse through A, B, C having the centroid of triangle ABC as center.

E 2991. *Proposed by Z. F. Starc, Vrsac, Yugoslavia.*

Prove or disprove the following inequality:

$$\alpha \tanh(x) > \sin(\alpha x) \quad (x > 0, \alpha > 1)$$

SOLUTIONS OF ELEMENTARY PROBLEMS

$f(n) = \gcd\{k^n - k | k = 2,3,4,\dots\}$

E 2901 [1981, 538]. *Proposed by S. C. Locke and A. Mandel, University of Waterloo.*

Let $f(n) = \gcd\{k^n - k | k = 2,3,4,\dots\}$ for $N \geqslant 2$. Evaluate $f(n)$. In particular, show that $f(2n) = 2$.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2989. *Proposed by M. Goldberg, University of Waterloo, and S. C. Locke, Florida Atlantic University.*

Let $A = (a, a')$ and $B = (b, b')$ be lattice points (points in the plane with integral coordinates), and let $d(A, B)$ denote $|a - b| + |a' - b'|$. Let S be the set of points at d-distance at most k from the origin. Calculate

$$f(k) = \sum_{A, B \in S} d(A, B).$$

E 2990. *Proposed by H. Eves, University of Maine, and C. Kimberling, University of Evansville.*

Let ABC be a triangle and L a line in the plane of ABC not passing through A, B, or C.

i) Prove that the isogonal conjugate of L is an ellipse, parabola or hyperbola according as L meets the circumcircle of ABC in zero, one or two points.
ii) Prove that the isotomic conjugate of L is an ellipse, parabola, or hyperbola according as L meets E in zero, one, or two points, where E is the ellipse through A, B, C having the centroid of triangle ABC as center.

E 2991. *Proposed by Z. F. Starc, Vrsac, Yugoslavia.*

Prove or disprove the following inequality:

$$\alpha \tanh(x) > \sin(\alpha x) \quad (x > 0, \alpha > 1)$$

SOLUTIONS OF ELEMENTARY PROBLEMS

$f(n) = \gcd\{k^n - k | k = 2,3,4,\dots\}$

E 2901 [1981, 538]. *Proposed by S. C. Locke and A. Mandel, University of Waterloo.*

Let $f(n) = \gcd\{k^n - k | k = 2,3,4,\dots\}$ for $N \geqslant 2$. Evaluate $f(n)$. In particular, show that $f(2n) = 2$.

Solution by Robert Breusch, Amherst College.

1. $f(n)$ cannot contain a factor p^2 for any prime p ($p \geqslant 2$) because $p^2 \nmid k(k^{n-1} - 1)$ for $k = p$.

2. For every n, $2 | f(n)$.

3. If p is an odd prime and if a is a primitive root (mod p), then $p | a(a^{n-1} - 1)$ only if $(p-1)|(n-1)$. On the other hand, if $(p-1)|(n-1)$ then $p|(k^n - k)$ for every k. Thus if P is the product of the distinct odd primes p such that $(p-1)|(n-1)$, then $f(n) = 2P$. (In particular $6|f(n)$ for every odd n.)

4. $(p-1)$ is not a divisor of $(2n-1)$ for any odd prime p. Thus $f(2n) = 2$.

Also solved by J. T. Arnold, L. Bělohrad (Czechoslovakia), M. Bencze (Romania), K. L. Bernstein, J. Browkin (Poland), P. S. Bruckman, H. M. Edgar, L. Foster, R. Gilmer, T. Herman, M. Josephy (Costa Rica), V. S. Joshi & A. M. Vaidya (India), W. T. M. Kars (Netherlands), R. E. Kennedy, J. O. Kiltinen, C. -N. Lee, M. Lehtinen (Finland), O. P. Lossers (Netherlands), J. G. Merickel (student), R. Morelli (student), R. B. Nelson, W. A. Newcomb, D. A. Rawsthorne, D. M. Rosenblum, R. P. Savage, Jr., R. E. Shafer, G. Shulman, S. Singh, D. Singmaster (England), D. B. Shapiro, L. Somer, J. M. Stark, A. Stenger, R. Teitler (United Kingdom), E. Trost (Switzerland), S. Wagon, M. Woltermann, P. Yiu (Canada), B. Zinner (Germany), an anonymous solver, and the proposer.

Kiltenen referred to his article, *Linearity of Exponentiation*, *Mathematics Magazine*, 52(1979) 3–9. Bruckman pointed out that $f(2m+1)$ is equal to the denominator of B_{2m}. Foster noted $f(n) = 6$ for infinitely many n.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1983. The solver's full post-office address should be on each sheet.

6422. *Proposed by Harvey Diamond, West Virginia University, and Bruce Montgomery, Carnegie-Mellon University.*

Let f: $[0, \infty) \to (-\infty, \infty)$ be a measurable function. If $p > 1$, is it possible that $f(x) - \int_0^x |f|^p \to \infty$ as $x \to \infty$?

SOLUTIONS OF ADVANCED PROBLEMS

Generation of the Alternating Group by Three Involutions

6350 [1981, 540]. *Proposed by James Wiegold, University College, Cardiff, Wales.*

Prove that the alternating group A_n ($n > 4$) can be generated by three involutions. Can two of them be conjugate? (*) Can one of them be preassigned?

Solution by Alan G. Williamson, University College, Galway, Ireland.

PROPOSITION. *For $n > 4$, the alternating group A_n can be generated by three involutions, one of which is preassigned and two of which are conjugate.*

Notation. We take A_n acting on the set $\Omega = \{1, 2, \ldots, n\}$. The *degree* of a permutation is the number of points that it displaces. Thus an involution in A_n has degree divisible by 4; and two involutions in A_n are conjugate (in A_n) if and only if they have the same degree. By relabelling the points (if necessary) we may assume that the preassigned involution is one of those defined below. Our construction of generators utilizes the following well-known result of Jordan:

THEOREM. (*See* [1].) *Let G be a doubly transitive permutation group of degree n. If either* (a) *G contains a 3-cycle, or* (b) *G contains an involution of degree 4 and $n > 8$, then G contains A_n.*

It is convenient to consider separately five cases. First of all, we deal with A_8:

Case A. $n = 8$. Let $u = (1,2)(3,4)$, $v = (1,5)(6,7)$, $w = (1,6)(3,7)(4,5)(2,8)$. Here $\langle u, v, w\rangle$ is transitive on Ω, $\langle u, v, w^{-1}vw\rangle$ is transitive on $\Omega \setminus \{8\}$. Hence $\langle u, v, w\rangle$ is doubly transitive on

Solution by Robert Breusch, Amherst College.

1. $f(n)$ cannot contain a factor p^2 for any prime p $(p \geqslant 2)$ because $p^2 \nmid k(k^{n-1}-1)$ for $k = p$.
2. For every n, $2|f(n)$.
3. If p is an odd prime and if a is a primitive root $(\bmod\, p)$, then $p|a(a^{n-1}-1)$ only if $(p-1)|(n-1)$. On the other hand, if $(p-1)|(n-1)$ then $p|(k^n - k)$ for every k. Thus if P is the product of the distinct odd primes p such that $(p-1)|(n-1)$, then $f(n) = 2P$. (In particular $6|f(n)$ for every odd n.)
4. $(p-1)$ is not a divisor of $(2n-1)$ for any odd prime p. Thus $f(2n) = 2$.

Also solved by J. T. Arnold, L. Bělohrad (Czechoslovakia), M. Bencze (Romania), K. L. Bernstein, J. Browkin (Poland), P. S. Bruckman, H. M. Edgar, L. Foster, R. Gilmer, T. Herman, M. Josephy (Costa Rica), V. S. Joshi & A. M. Vaidya (India), W. T. M. Kars (Netherlands), R. E. Kennedy, J. O. Kiltinen, C. -N. Lee, M. Lehtinen (Finland), O. P. Lossers (Netherlands), J. G. Merickel (student), R. Morelli (student), R. B. Nelson, W. A. Newcomb, D. A. Rawsthorne, D. M. Rosenblum, R. P. Savage, Jr., R. E. Shafer, G. Shulman, S. Singh, D. Singmaster (England), D. B. Shapiro, L. Somer, J. M. Stark, A. Stenger, R. Teitler (United Kingdom), E. Trost (Switzerland), S. Wagon, M. Woltermann, P. Yiu (Canada), B. Zinner (Germany), an anonymous solver, and the proposer.

Kiltenen referred to his article, *Linearity of Exponentiation*, *Mathematics Magazine*, 52(1979) 3–9. Bruckman pointed out that $f(2m+1)$ is equal to the denominator of B_{2m}. Foster noted $f(n) = 6$ for infinitely many n.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1983. The solver's full post-office address should be on each sheet.

6422. *Proposed by Harvey Diamond, West Virginia University, and Bruce Montgomery, Carnegie-Mellon University.*

Let f: $[0,\infty) \to (-\infty,\infty)$ be a measurable function. If $p > 1$, is it possible that $f(x) - \int_0^x |f|^p \to \infty$ as $x \to \infty$?

SOLUTIONS OF ADVANCED PROBLEMS

Generation of the Alternating Group by Three Involutions

6350 [1981, 540]. *Proposed by James Wiegold, University College, Cardiff, Wales.*

Prove that the alternating group A_n $(n > 4)$ can be generated by three involutions. Can two of them be conjugate? (*) Can one of them be preassigned?

Solution by Alan G. Williamson, University College, Galway, Ireland.

PROPOSITION. *For $n > 4$, the alternating group A_n can be generated by three involutions, one of which is preassigned and two of which are conjugate.*

Notation. We take A_n acting on the set $\Omega = \{1, 2, \ldots, n\}$. The *degree* of a permutation is the number of points that it displaces. Thus an involution in A_n has degree divisible by 4; and two involutions in A_n are conjugate (in A_n) if and only if they have the same degree. By relabelling the points (if necessary) we may assume that the preassigned involution is one of those defined below. Our construction of generators utilizes the following well-known result of Jordan:

THEOREM. (*See* [1].) *Let G be a doubly transitive permutation group of degree n. If either* (a) *G contains a 3-cycle, or* (b) *G contains an involution of degree 4 and $n > 8$, then G contains A_n.*

It is convenient to consider separately five cases. First of all, we deal with A_8:

Case A. $n = 8$. Let $u = (1,2)(3,4)$, $v = (1,5)(6,7)$, $w = (1,6)(3,7)(4,5)(2,8)$. Here $\langle u, v, w\rangle$ is transitive on Ω, $\langle u, v, w^{-1}vw\rangle$ is transitive on $\Omega \setminus \{8\}$. Hence $\langle u, v, w\rangle$ is doubly transitive on

Solution by Robert Breusch, Amherst College.

1. $f(n)$ cannot contain a factor p^2 for any prime p $(p \geqslant 2)$ because $p^2 \nmid k(k^{n-1} - 1)$ for $k = p$.
2. For every n, $2 | f(n)$.
3. If p is an odd prime and if a is a primitive root (mod p), then $p | a(a^{n-1} - 1)$ only if $(p - 1) | (n - 1)$. On the other hand, if $(p - 1) | (n - 1)$ then $p | (k^n - k)$ for every k. Thus if P is the product of the distinct odd primes p such that $(p - 1) | (n - 1)$, then $f(n) = 2P$. (In particular $6 | f(n)$ for every odd n.)
4. $(p - 1)$ is not a divisor of $(2n - 1)$ for any odd prime p. Thus $f(2n) = 2$.

Also solved by J. T. Arnold, L. Bĕlohrad (Czechoslovakia), M. Bencze (Romania), K. L. Bernstein, J. Browkin (Poland), P. S. Bruckman, H. M. Edgar, L. Foster, R. Gilmer, T. Herman, M. Josephy (Costa Rica), V. S. Joshi & A. M. Vaidya (India), W. T. M. Kars (Netherlands), R. E. Kennedy, J. O. Kiltinen, C. -N. Lee, M. Lehtinen (Finland), O. P. Lossers (Netherlands), J. G. Merickel (student), R. Morelli (student), R. B. Nelson, W. A. Newcomb, D. A. Rawsthorne, D. M. Rosenblum, R. P. Savage, Jr., R. E. Shafer, G. Shulman, S. Singh, D. Singmaster (England), D. B. Shapiro, L. Somer, J. M. Stark, A. Stenger, R. Teitler (United Kingdom), E. Trost (Switzerland), S. Wagon, M. Woltermann, P. Yiu (Canada), B. Zinner (Germany), an anonymous solver, and the proposer.

Kiltenen referred to his article, *Linearity of Exponentiation*, *Mathematics Magazine*, 52(1979) 3–9. Bruckman pointed out that $f(2m + 1)$ is equal to the denominator of B_{2m}. Foster noted $f(n) = 6$ for infinitely many n.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1983. The solver's full post-office address should be on each sheet.

6422. *Proposed by Harvey Diamond, West Virginia University, and Bruce Montgomery, Carnegie-Mellon University.*

Let f: $[0, \infty) \to (-\infty, \infty)$ be a measurable function. If $p > 1$, is it possible that $f(x) - \int_0^x |f|^p \to \infty$ as $x \to \infty$?

SOLUTIONS OF ADVANCED PROBLEMS

Generation of the Alternating Group by Three Involutions

6350 [1981, 540]. *Proposed by James Wiegold, University College, Cardiff, Wales.*

Prove that the alternating group A_n $(n > 4)$ can be generated by three involutions. Can two of them be conjugate? (*) Can one of them be preassigned?

Solution by Alan G. Williamson, University College, Galway, Ireland.

PROPOSITION. *For $n > 4$, the alternating group A_n can be generated by three involutions, one of which is preassigned and two of which are conjugate.*

Notation. We take A_n acting on the set $\Omega = \{1, 2, \ldots, n\}$. The *degree* of a permutation is the number of points that it displaces. Thus an involution in A_n has degree divisible by 4; and two involutions in A_n are conjugate (in A_n) if and only if they have the same degree. By relabelling the points (if necessary) we may assume that the preassigned involution is one of those defined below. Our construction of generators utilizes the following well-known result of Jordan:

THEOREM. (*See* [1].) *Let G be a doubly transitive permutation group of degree n. If either (a) G contains a 3-cycle, or (b) G contains an involution of degree 4 and $n > 8$, then G contains A_n.*

It is convenient to consider separately five cases. First of all, we deal with A_8:

Case A. $n = 8$. Let $u = (1,2)(3,4)$, $v = (1,5)(6,7)$, $w = (1,6)(3,7)(4,5)(2,8)$. Here $\langle u, v, w\rangle$ is transitive on Ω, $\langle u, v, w^{-1}vw\rangle$ is transitive on $\Omega \setminus \{8\}$. Hence $\langle u, v, w\rangle$ is doubly transitive on

Ω and contains the 3-cycle $(uv)^2$.

Case B. $n = 4m > 8$. Let

$$x = (1,2)\,(3,4)\cdots(2r-1,2r)\cdots(4m-1,4m)$$

of degree 4m. Let

$$y = (1,2)\,(4,5)\cdots(2r,2r+1)\cdots(4m-2,4m-1)\,(3,4m)$$

of degree $4m$. For any positive integer $s < m$, we define

$$z = (1,3)\,(4,5)\cdots(2r,2r+1)\cdots(4s,4s+1)$$

of degree $4s$. Since $xy = (3,5,\ldots,4m-1)\,(4m,4m-2,\ldots,4)$, it is clear that $\langle xy, z\rangle$ is transitive on $\Omega\setminus\{2\}$ and $\langle x, y, z\rangle$ is transitive on Ω. Therefore $\langle x, y, z\rangle$ is doubly transitive on Ω and contains the involution $(yz)^2 = (1,3)\,(2,4m)$.

Case C. $n = 4m + 1 \geqslant 5$. Take x as above. For $m > 1$, we define

$$v = (2,3)\,(4,5)\cdots(2r,2r+1)\cdots(4m-2,4m-1)\,(1,4m)(4m+1)$$

of degree $4m$. (For $m = 1$, take $v = (2,3)\,(4,5)$.) Let

$$w = (4m-4s+3,4m-4s+4)\cdots(2r-1,2r)\cdots(4m-1,4m)\,(1,4m+1)$$

of degree $4s$. Here $\langle x, v\rangle$ is transitive on $\Omega\setminus\{4m+1\}$, $\langle x, v, w\rangle$ is transitive on Ω. So $\langle x, v, w\rangle$ is doubly transitive on Ω, and contains the 3-cycle $(xw)^2 = (1,4m+1,2)$.

Case D. $n = 4m + 2 \geqslant 6$. Take x as above. For $m > 1$, we define

$$u = (1,3)\,(2,5)\cdots(2r,2r+3)\cdots(4m-4,4m-1)\,(4m-2,4m+1)\,(4m)\,(4m+2)$$

of degree $4m$. (For $m = 1$, take $u = (1,3)\,(2,5)$.) Then for $s > 1$, we define

$$t = (4m-4s+1,4m-4s+2)\cdots(2r-1,2r)\cdots$$
$$\times(4m-3,4m-2)\,(4m-1,4m+2)\,(4m)\,(4m+1)$$

of degree $4s$. (For $s = 1$, we take $t = (4m-3,4m-2)\,(4m-1,4m+2)$ of degree 4.) Now $\langle x, u\rangle$ is transitive on $\Omega\setminus\{4m+2\}$, $\langle x, u, t\rangle$ is transitive on Ω. So $\langle x, u, t\rangle$ is doubly transitive on Ω and contains the 3-cycle $(xt)^2 = (4m,4m-1,4m+2)$.

Case E. $n = 4m + 3 \geqslant 7$. Take x and u as above. For $m > s > 1$, we define

$$d = (4m-4s+1,4m-4s+2)\cdots(2r-1,2r)\cdots$$
$$\times(4m-5,4m-4)\,(4m+1,4m+2)\,(4m-1,4m+3),$$

which is of degree $4s$ and fixes the points $4m, 4m-2, 4m-3$. (For $s = 1$, take $d = (4m+1,4m+2)\,(4m-1,4m+3)$.) Here $\langle x, u\rangle$ is transitive on $\Omega\setminus\{4m+2,4m+3\}$ and $\langle x, u, d\rangle$ is transitive on Ω. Now $\langle x, u, d^{-1}xd\rangle$ is transitive on $\Omega\setminus\{4m+2\}$, because $d^{-1}xd$ interchanges $4m$ and $4m+3$, whilst fixing $4m+2$. Hence $\langle x, u, d\rangle$ is doubly transitive and contains the 3-cycle $(xd)^2 = (4m,4m-1,4m+3)$.

Reference

1. H. Wielandt, Finite permutation groups, Academic Press, New York, 1964.

Also solved by LeRoy Beasley, who refers to a theorem of Alan Williamson: *If a primitive transitive permutation group G of degree n contains a t-cycle, then G contains A_n unless $t > (n-t)!$.*

Perron Roots

6354 [1981, 541] *Proposed by Emeric Deutsch*, Polytechnic Institute of New York.

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that there exists a

Ω and contains the 3-cycle $(uv)^2$.

Case B. $n = 4m > 8$. Let

$$x = (1,2)\,(3,4)\cdots(2r-1,2r)\cdots(4m-1,4m)$$

of degree 4m. Let

$$y = (1,2)\,(4,5)\cdots(2r,2r+1)\cdots(4m-2,4m-1)\,(3,4m)$$

of degree $4m$. For any positive integer $s < m$, we define

$$z = (1,3)\,(4,5)\cdots(2r,2r+1)\cdots(4s,4s+1)$$

of degree $4s$. Since $xy = (3,5,\ldots,4m-1)\,(4m,4m-2,\ldots,4)$, it is clear that $\langle xy, z\rangle$ is transitive on $\Omega\setminus\{2\}$ and $\langle x, y, z\rangle$ is transitive on Ω. Therefore $\langle x, y, z\rangle$ is doubly transitive on Ω and contains the involution $(yz)^2 = (1,3)\,(2,4m)$.

Case C. $n = 4m+1 \geqslant 5$. Take x as above. For $m > 1$, we define

$$v = (2,3)\,(4,5)\cdots(2r,2r+1)\cdots(4m-2,4m-1)\,(1,4m)(4m+1)$$

of degree $4m$. (For $m = 1$, take $v = (2,3)\,(4,5)$.) Let

$$w = (4m-4s+3,4m-4s+4)\cdots(2r-1,2r)\cdots(4m-1,4m)\,(1,4m+1)$$

of degree $4s$. Here $\langle x, v\rangle$ is transitive on $\Omega\setminus\{4m+1\}$, $\langle x, v, w\rangle$ is transitive on Ω. So $\langle x, v, w\rangle$ is doubly transitive on Ω, and contains the 3-cycle $(xw)^2 = (1,4m+1,2)$.

Case D. $n = 4m+2 \geqslant 6$. Take x as above. For $m > 1$, we define

$$u = (1,3)\,(2,5)\cdots(2r,2r+3)\cdots(4m-4,4m-1)\,(4m-2,4m+1)\,(4m)\,(4m+2)$$

of degree $4m$. (For $m = 1$, take $u = (1,3)\,(2,5)$.) Then for $s > 1$, we define

$$t = (4m-4s+1,4m-4s+2)\cdots(2r-1,2r)\cdots$$
$$\times(4m-3,4m-2)\,(4m-1,4m+2)\,(4m)\,(4m+1)$$

of degree $4s$. (For $s = 1$, we take $t = (4m-3,4m-2)\,(4m-1,4m+2)$ of degree 4.) Now $\langle x, u\rangle$ is transitive on $\Omega\setminus\{4m+2\}$, $\langle x, u, t\rangle$ is transitive on Ω. So $\langle x, u, t\rangle$ is doubly transitive on Ω and contains the 3-cycle $(xt)^2 = (4m,4m-1,4m+2)$.

Case E. $n = 4m+3 \geqslant 7$. Take x and u as above. For $m > s > 1$, we define

$$d = (4m-4s+1,4m-4s+2)\cdots(2r-1,2r)\cdots$$
$$\times(4m-5,4m-4)\,(4m+1,4m+2)\,(4m-1,4m+3),$$

which is of degree $4s$ and fixes the points $4m, 4m-2, 4m-3$. (For $s = 1$, take $d = (4m+1,4m+2)\,(4m-1,4m+3)$.) Here $\langle x, u\rangle$ is transitive on $\Omega\setminus\{4m+2,4m+3\}$ and $\langle x, u, d\rangle$ is transitive on Ω. Now $\langle x, u, d^{-1}xd\rangle$ is transitive on $\Omega\setminus\{4m+2\}$, because $d^{-1}xd$ interchanges $4m$ and $4m+3$, whilst fixing $4m+2$. Hence $\langle x, u, d\rangle$ is doubly transitive and contains the 3-cycle $(xd)^2 = (4m,4m-1,4m+3)$.

Reference

1. H. Wielandt, Finite permutation groups, Academic Press, New York, 1964.

Also solved by LeRoy Beasley, who refers to a theorem of Alan Williamson: *If a primitive transitive permutation group G of degree n contains a t-cycle, then G contains A_n unless $t > (n-t)!$.*

Perron Roots

6354 [1981, 541] *Proposed by Emeric Deutsch*, Polytechnic Institute of New York.

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that there exists a

constant $K > 0$ such that

$$\|(tI - A)^{-1}\| \leqslant \frac{K}{t - r} \text{ for all } t > r.$$

Here I denotes the identity matrix and $\|\ \ \|$ is the operator norm induced by the Euclidean norm $x \to (x^T x)^{1/2}$. *Find the best value of the constant K.

Solution by H. Kestelman, University College London. We have to show that $\|M_t\| \leqslant K$ for all $t > r$, where $M_t = (t - r)(tI - A)^{-1}$. It is well known that $M_t \geqslant 0$. From the hypothesis on A, there is a positive vector v satisfying $Av = rv$ and therefore $M_t v = v$. If u is the vector with all components 1, then $pu \leqslant v \leqslant qu$ where p is the least and q the greatest component of v. Now take any x with $\|x\| = 1$; then $|x| \leqslant u$ and

$$0 \leqslant M_t|x| \leqslant M_t u \leqslant p^{-1}M_t v = p^{-1}v \leqslant qp^{-1}u.$$

So

$$\|M_t\| \leqslant qp^{-1}\|u\| = qp^{-1}n^{1/2}.$$

Also solved by Michele Elia (Italy), Mauri Koskela (Finland), Beresford Parlett, and the proposer.

No expression for the best value of K substantially better than the trivial $K = \sup_{t>r}\|M_t\|$ was obtained. There is no bound on K as A varies. To see this consider $A = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}$, $a > 0$. Then $r = a^{1/2}$, $M_t = (t + r)^{-1}(tI + A)$ and for $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $t > r$, we have

$$\|M_t\| \geqslant \|M_t x\| = (t + r)^{-1}(t^2 + 1)^{1/2} \text{ so that } K \geqslant \frac{1}{2}\left(1 + \frac{1}{a}\right)^{1/2}.$$

Abundant and Deficient Numbers

6356 [1981, 623]. *Proposed by C. R. Wall, Trident Technical College, Charleston, S.C.*

A number N is called deficient if the sum of its proper divisors is less than N: $\sigma(N) < 2N$. It is called abundant if $\sigma(N) > 2N$.

(a) Let k be fixed. Do there exist sequences of k consecutive abundant numbers?

(b) Prove that there are infinitely many sequences of 5 consecutive deficient numbers.

Solution by Lorraine L. Foster, California State University, Northridge, California. Let p_n denote the nth prime.

(a) Since $\prod(1 + 1/p_n) = \infty$, for each positive integer m there is an integer $N(m)$ such that $\prod_{n=m}^{N(m)} (1 + 1/p_n) > 2$. Hence $\prod_{n=m}^{N(m)} p_n$ is abundant. Define $m_1 = 1$; $m_{j+1} = N(m_j) + 1$ for $j \geqslant 1$, and let

$$t_j = \prod_{n=m_j}^{N(m_j)} p_n \text{ for } j \geqslant 1.$$

Since the t_j are pairwise relatively prime, we may choose $X \geqslant 0$ such that $X + j \equiv 0 \pmod{t_j}$ for $j = 1, 2, \ldots, k$. Since a multiple of an abundant number is abundant, $X + 1, X + 2, \ldots, X + k$ is a sequence of k consecutive abundant numbers.

(b) Since $\lim_{n\to\infty} (1 + 2/(n \log n))^n = 1$ and, by the prime number theorem, $\lim_{n\to\infty} p_n/(n \log n) = 1$, we may suppose n to be so large that

$$(1 + 2/(n \log n))^n < 1.01 \text{ and } p_n > (n \log n)/2 > 60.$$

Define $P_n = \prod_{j=1}^{n-1} p_j$, $M_n = 60P_n + 1$, and let $x_i = (M_n + i - 1)/i$ for $i = 1, 2, 3, 4, 5$. Since $n > 3$, it is easy to see that each x_i is an integer relatively prime to P_n. Hence each x_i is of the form $\prod_{j=1}^{t} P_{n_j}$ where $t \leqslant n$, $n_j \geqslant n$ (and the n_j are not necessarily distinct). Since $1 + 1/p_{n_j} \leqslant 1 + p_n$

constant $K > 0$ such that

$$\|(tI - A)^{-1}\| \leqslant \frac{K}{t-r} \text{ for all } t > r.$$

Here I denotes the identity matrix and $\| \quad \|$ is the operator norm induced by the Euclidean norm $x \to (x^T x)^{1/2}$. *Find the best value of the constant K.

Solution by H. Kestelman, University College London. We have to show that $\|M_t\| \leqslant K$ for all $t > r$, where $M_t = (t-r)(tI - A)^{-1}$. It is well known that $M_t \geqslant 0$. From the hypothesis on A, there is a positive vector v satisfying $Av = rv$ and therefore $M_t v = v$. If u is the vector with all components 1, then $pu \leqslant v \leqslant qu$ where p is the least and q the greatest component of v. Now take any x with $\|x\| = 1$; then $|x| \leqslant u$ and

$$0 \leqslant M_t|x| \leqslant M_t u \leqslant p^{-1}M_t v = p^{-1}v \leqslant qp^{-1}u.$$

So

$$\|M_t\| \leqslant qp^{-1}\|u\| = qp^{-1}n^{1/2}.$$

Also solved by Michele Elia (Italy), Mauri Koskela (Finland), Beresford Parlett, and the proposer.

No expression for the best value of K substantially better than the trivial $K = \sup_{t>r}\|M_t\|$ was obtained. There is no bound on K as A varies. To see this consider $A = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix}$, $a > 0$. Then $r = a^{1/2}$, $M_t = (t+r)^{-1}(tI + A)$ and for $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $t > r$, we have

$$\|M_t\| \geqslant \|M_t x\| = (t+r)^{-1}(t^2+1)^{1/2} \text{ so that } K \geqslant \frac{1}{2}\left(1 + \frac{1}{a}\right)^{1/2}.$$

Abundant and Deficient Numbers

6356 [1981, 623]. *Proposed by C. R. Wall, Trident Technical College, Charleston, S.C.*

A number N is called deficient if the sum of its proper divisors is less than N: $\sigma(N) < 2N$. It is called abundant if $\sigma(N) > 2N$.

(a) Let k be fixed. Do there exist sequences of k consecutive abundant numbers?

(b) Prove that there are infinitely many sequences of 5 consecutive deficient numbers.

Solution by Lorraine L. Foster, California State University, Northridge, California. Let p_n denote the nth prime.

(a) Since $\prod(1 + 1/p_n) = \infty$, for each positive integer m there is an integer $N(m)$ such that $\prod_{n=m}^{N(m)}(1 + 1/p_n) > 2$. Hence $\prod_{n=m}^{N(m)} p_n$ is abundant. Define $m_1 = 1$; $m_{j+1} = N(m_j) + 1$ for $j \geqslant 1$, and let

$$t_j = \prod_{n=m_j}^{N(m_j)} p_n \text{ for } j \geqslant 1.$$

Since the t_j are pairwise relatively prime, we may choose $X \geqslant 0$ such that $X + j \equiv 0 \pmod{t_j}$ for $j = 1, 2, \ldots, k$. Since a multiple of an abundant number is abundant, $X+1, X+2, \ldots, X+k$ is a sequence of k consecutive abundant numbers.

(b) Since $\lim_{n\to\infty}(1 + 2/(n\log n))^n = 1$ and, by the prime number theorem, $\lim_{n\to\infty} p_n/(n\log n) = 1$, we may suppose n to be so large that

$$(1 + 2/(n\log n))^n < 1.01 \text{ and } p_n > (n\log n)/2 > 60.$$

Define $P_n = \prod_{j=1}^{n-1} p_j$, $M_n = 60P_n + 1$, and let $x_i = (M_n + i - 1)/i$ for $i = 1, 2, 3, 4, 5$. Since $n > 3$, it is easy to see that each x_i is an integer relatively prime to P_n. Hence each x_i is of the form $\prod_{j=1}^{t} P_{n_j}$ where $t \leqslant n$, $n_j \geqslant n$ (and the n_j are not necessarily distinct). Since $1 + 1/p_{n_j} \leqslant 1 + p_n$

and $\sigma(p^s)/p^s \leqslant (\sigma(p)/p)^s$ for $s \geqslant 1$, p prime, we conclude that

$$\sigma(x_i)/x_i \leqslant (1 + 1/p_n)^n < (1 + 2/(n\log n))^n < 1.01 \text{ for } i = 1,2,3,4,5.$$

Thus

$$\frac{\sigma(M_n + i - 1)}{M_n + i} = \frac{\sigma(i)}{i}\frac{\sigma(x_i)}{x_i} < (1.75)(1.01) < 2 \text{ for } i = 1,2,3,4,5,$$

and so $M_n, M_n + 1, \ldots, M_n + 4$ is a sequence of 5 consecutive deficient numbers. The result is best possible since any multiple of 6 is abundant.

Also solved by Claudia Spiro and the proposer. Jerzy Browkin (Poland) and Jeff Loveland also solved part (a).

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Mathematics: *Problem Solving through Recreational Mathematics*. By Bonnie Averbach and Orin Chein. W. H. Freeman, San Francisco, 1980. pp. vii + 400. $16.50. Accompanying instructor's guide, 169 pp.

MURRAY S. KLAMKIN
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G2G1

For some time now, Problem Solving has been in high fashion in mathematical education circles. As a consequence, there has been a steady stream of books in this field. Whether or not you judge these books to be worthwhile depends highly on your own background and your own views on Problem Solving. Also, one should take into consideration the aims of the authors even if it turns out that one is not wholly in agreement with them.

What is Problem Solving? As an indication, Halmos in a rather charming article [**4**] notes that:

> Mathematicians sometimes classify themselves and each other as either problem-solvers or theory-creators. The problem-solvers answer yes-or-no questions and discuss vital special cases and concrete examples that are the flesh and blood of all mathematics; the theory-creators fit the results into a framework, illuminate it all, and point it in a definite direction—they provide the skeleton and the soul of mathematics. One and the same human being can be both a problem-solver and a theory-creator, but usually, he is mainly one or the other. The problem-solvers make geometric constructions, the theory-creators discuss the foundations of Euclidean geometry; the problem-solvers find out what makes switching diagrams tick, the theory-creators prove representation theorems for Boolean algebras.

Many problem-solvers (and I classify myself that way) are not only concerned with solving problems per se but also with connecting them with other problems, and extending their applicability, i.e., theory-creating. Dually, the theory-creators have to be problem-solvers in establishing their necessary sequences of lemmas and theorems for their theories.

As a better indication for Problem Solving, I can do no better than to quote from George Pólya. His five books [**5**]–[**9**] are still, in my view, the best books written on Problem Solving.

> A great discovery solves a great problem but there is a grain of discovery in the solution of any problem. Your problem may be modest; but if it challenges your curiosity and brings into play your inventive faculties, and if you solve it by your own means, you may experience the tension and enjoy

and $\sigma(p^s)/p^s \leqslant (\sigma(p)/p)^s$ for $s \geqslant 1$, p prime, we conclude that

$$\sigma(x_i)/x_i \leqslant (1 + 1/p_n)^n < (1 + 2/(n\log n))^n < 1.01 \text{ for } i = 1,2,3,4,5.$$

Thus

$$\frac{\sigma(M_n + i - 1)}{M_n + i} = \frac{\sigma(i)}{i}\frac{\sigma(x_i)}{x_i} < (1.75)(1.01) < 2 \text{ for } i = 1,2,3,4,5,$$

and so $M_n, M_n + 1, \ldots, M_n + 4$ is a sequence of 5 consecutive deficient numbers. The result is best possible since any multiple of 6 is abundant.

Also solved by Claudia Spiro and the proposer. Jerzy Browkin (Poland) and Jeff Loveland also solved part (a).

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Mathematics: Problem Solving through Recreational Mathematics. By Bonnie Averbach and Orin Chein. W. H. Freeman, San Francisco, 1980. pp. vii + 400. $16.50. Accompanying instructor's guide, 169 pp.

MURRAY S. KLAMKIN
Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G2G1

For some time now, Problem Solving has been in high fashion in mathematical education circles. As a consequence, there has been a steady stream of books in this field. Whether or not you judge these books to be worthwhile depends highly on your own background and your own views on Problem Solving. Also, one should take into consideration the aims of the authors even if it turns out that one is not wholly in agreement with them.

What is Problem Solving? As an indication, Halmos in a rather charming article [4] notes that:

> Mathematicians sometimes classify themselves and each other as either problem-solvers or theory-creators. The problem-solvers answer yes-or-no questions and discuss vital special cases and concrete examples that are the flesh and blood of all mathematics; the theory-creators fit the results into a framework, illuminate it all, and point it in a definite direction—they provide the skeleton and the soul of mathematics. One and the same human being can be both a problem-solver and a theory-creator, but usually, he is mainly one or the other. The problem-solvers make geometric constructions, the theory-creators discuss the foundations of Euclidean geometry; the problem-solvers find out what makes switching diagrams tick, the theory-creators prove representation theorems for Boolean algebras.

Many problem-solvers (and I classify myself that way) are not only concerned with solving problems per se but also with connecting them with other problems, and extending their applicability, i.e., theory-creating. Dually, the theory-creators have to be problem-solvers in establishing their necessary sequences of lemmas and theorems for their theories.

As a better indication for Problem Solving, I can do no better than to quote from George Pólya. His five books [5]–[9] are still, in my view, the best books written on Problem Solving.

> A great discovery solves a great problem but there is a grain of discovery in the solution of any problem. Your problem may be modest; but if it challenges your curiosity and brings into play your inventive faculties, and if you solve it by your own means, you may experience the tension and enjoy

the triumph of discovery. Such experiences at a susceptible age may create a taste for mental work and leave their imprint on mind and character for a lifetime.

Thus, a teacher of mathematics has a great opportunity. If he fills his allotted time with drilling his students in routine operations he kills their interest, hampers their intellectual development, and misuses his opportunity. But if he challenges the curiosity of his students by setting them problems proportionate to their knowledge, and helps them to solve their problems with stimulating questions, he may give them a taste for, and some means of, independent thinking. [**5**, p.v]

1. Solving a problem means finding a way out of a difficulty, a way around an obstacle, attaining an aim which was not immediately attainable. Solving problems is the specific achievement of intelligence, and intelligence is the specific gift of mankind: solving problems can be regarded as the most characteristically human activity. The aim of this work is to understand this activity, to propose means to teach it, and, eventually, to improve the problem-solving ability of the reader.

2. This work consists of two parts; let me characterize briefly the role of these two parts.

Solving problems is a practical art, like swimming, or skiing, or playing the piano: you can learn it only by imitation and practise. This book cannot offer you a magic key that opens all the doors and solves all the problems, but it offers you good examples for imitation and many opportunities for practise: if you wish to become a problem solver you have to solve problems. [**6**, p.v]

The authors, Averbach and Chein, address the nontrivial problem of how liberal arts students can be given a feeling for mathematics in a one or two semester college course. They have concluded that with most of the usual approaches through history, culture, or applications, the student is merely an observer which is not good enough. And in line with Pólya's comments, their aim is to get the students to think for themselves on problems which are not essentially the same as those presented in class. They eschew applications since they feel that only highly simplified cases can usually be considered. Their approach is based on recreational mathematics—recreational problems, puzzles and games, since firstly they are fun and provide motivation which clearly is important and, secondly, historically many important mathematical problems arose from problems which were recreational in origin. While I agree with these two latter points, it should also be noted that many recreational problems were not important mathematically and that many important mathematical problems arose from considering applications. Also, the authors' emphasis is not on the mathematical results themselves but rather on how these results can be used in thinking about problems and solving them.

If one accepts the authors' premises, then they have done a good job and it is likely that students using the book will be motivated by the discussions and problems. The chapters are essentially independent and each contains several types of problems. Many of them have appeared previously in columns, journals, and puzzle or problem books in recreational mathematics. There are sample problems with solutions to whet the reader's appetite and to motivate the discussions in the chapters. As a useful aid to the student, the exercises are marked in three ways; those that are solved completely in the Hints and Solution section, those that have answers provided and those that provide helpful hints. Problems are also prefixed with one or two stars to denote those problems which are difficult or lengthy, or very difficult, respectively.

I found myself being somewhat turned off by the book, after an initial refreshing feeling, because of the overemphasis on the recreational aspects. No doubt, part of this was due to overexposure since I had already seen very many of the problems given or very similar ones. I would prefer a book for liberal arts students which blended in some history and culture together with problem solving, both pure and applied. Mathematics is much more than just fun and games, and to me the authors have overemphasized these aspects. Also, I do not agree with the authors that there are not any relatively simple realistic applications which can be treated. For example, one may consider the principles behind the following three instruments: a vernier, a stroboscope, and a spherometer. The first two lead to nice arithmetic applications while the third one is geometric. As a precursor for the third one, one can pose the problem of giving a Euclidean construction for the diameter of a given solid ball using a compass, a straight edge and a plain plane piece of paper. For other elementary applications, see [**11**], [**13**].

Other books which can be compared to the present one for the same purpose are [**1**], [**2**], [**3**], [**10**], [**14**]. Whether they are actually better or not cannot really be decided upon unless we also receive input from students who actually use these books for the purposes intended.

Finally, if I had to choose a text to be used solely for problem solving, I would prefer picking one or two of the five mentioned books by Pólya, since they have a better treatment of "Problem Solving Strategies," such as symmetry, continuity, physical analogies, etc., and are beautifully written. Nevertheless, at the least, the Averbach-Chein book is definitely useful as a supplementary text or reference since it does provide much interesting material for the students.

References

1. D. M. Campbell, The Whole Craft of Number, Prindle, Weber, and Schmidt, Boston, 1976.

2. H. Fremont, Teaching Secondary Mathematics through Applications, Prindle, Weber, and Schmidt, Boston, 1979.

3. S. Gudder, A Mathematical Journey, McGraw-Hill, New York, 1976.

4. P. R. Halmos, Mathematics as a creative art, Amer. Sci. 56 (1968) 375–389.

5. G. Pólya, How to Solve It, Doubleday Anchor Books, New York, 1957.

6. ______, Mathematical Discovery I, Wiley, New York, 1962.

7. ______, Mathematical Discovery II, Wiley, New York, 1965.

8. ______, Mathematics and Plausible Reasoning I, Princeton University Press, Princeton, New Jersey, 1954.

9. ______, Mathematics and Plausible Reasoning II, Princeton University Press, Princeton, New Jersey, 1954.

10. A. W. Roberts and D. L. Varberg, Faces of Mathematics, T. Y. Crowell, New York, 1978.

11. S. Sharon (Editor), Applications in School Mathematics, N.C.T.M., Reston, Virginia, 1979.

12. K. J. Smith, The Nature of Modern Mathematics, Brooks-Cole, Monterey, California, 1976.

13. A Source Book of Applications of School Mathematics, N.C.T.M., Reston, Virginia, 1980.

14. S. K. Stein, Mathematics, The Man-Made Universe, W. H. Freeman, San Francisco, 1976.

Quantum Mechanics in Hilbert Space, Second Edition. By Eduard Prugovecki. Academic Press, New York, 1981. xxii + 685 pp., $39.50.

JOHN CHALLIFOUR
Department of Mathematics, Indiana University, Bloomington, IN 47405

The power and beauty of mathematical models in physics has often been noted, sometimes with frustration but most often with awe. Quantum mechanics is no exception, and has come to mean those mathematical models or equations which describe the dynamics of the atomic and molecular world. At the mathematical level quantum mechanics describes physical phenomena by assigning to each "idealized state" of a physical system a unit vector in a complex (separable) Hilbert space. Physical measurements, or observables, are related to appropriate self-adjoint operators in Hilbert space, and the actual measured values are directly associated with the spectra of the corresponding operators. Typically, such a unit vector might represent the location of the valence electrons in an atom, and a positive, self-adjoint operator, called the Hamiltonian, has as its spectrum the energy levels which appear as spectral lines in the laboratory.

Since its inception in 1925, nonrelativistic quantum mechanics, independently due to W. Heisenberg and E. Schrödinger, has occupied a central place in the study of self-adjoint operators in Hilbert space by providing nontrivial problems that challenge the usefulness of the general theory and, by their physical setting, offer hints for solution. Among these are the spectral theorem (E. Hellinger, 1909; J. von Neumann, 1929), perturbations of unbounded operators (F. Rellich, 1939; T. Kato, 1957), wave operators and perturbation of absolutely continuous spectra (C. Moller, 1946; K. Friedrichs, 1948; T. Kato and S. T. Kuroda, 1959), and a general theory of operator algebras in Hilbert space (J. von Neumann, 1927; F. J. Murray and J. von

Other books which can be compared to the present one for the same purpose are [**1**], [**2**], [**3**], [**10**], [**14**]. Whether they are actually better or not cannot really be decided upon unless we also receive input from students who actually use these books for the purposes intended.

Finally, if I had to choose a text to be used solely for problem solving, I would prefer picking one or two of the five mentioned books by Pólya, since they have a better treatment of "Problem Solving Strategies," such as symmetry, continuity, physical analogies, etc., and are beautifully written. Nevertheless, at the least, the Averbach-Chein book is definitely useful as a supplementary text or reference since it does provide much interesting material for the students.

References

1. D. M. Campbell, The Whole Craft of Number, Prindle, Weber, and Schmidt, Boston, 1976.

2. H. Fremont, Teaching Secondary Mathematics through Applications, Prindle, Weber, and Schmidt, Boston, 1979.

3. S. Gudder, A Mathematical Journey, McGraw-Hill, New York, 1976.

4. P. R. Halmos, Mathematics as a creative art, Amer. Sci. 56 (1968) 375–389.

5. G. Pólya, How to Solve It, Doubleday Anchor Books, New York, 1957.

6. ______, Mathematical Discovery I, Wiley, New York, 1962.

7. ______, Mathematical Discovery II, Wiley, New York, 1965.

8. ______, Mathematics and Plausible Reasoning I, Princeton University Press, Princeton, New Jersey, 1954.

9. ______, Mathematics and Plausible Reasoning II, Princeton University Press, Princeton, New Jersey, 1954.

10. A. W. Roberts and D. L. Varberg, Faces of Mathematics, T. Y. Crowell, New York, 1978.

11. S. Sharon (Editor), Applications in School Mathematics, N.C.T.M., Reston, Virginia, 1979.

12. K. J. Smith, The Nature of Modern Mathematics, Brooks-Cole, Monterey, California, 1976.

13. A Source Book of Applications of School Mathematics, N.C.T.M., Reston, Virginia, 1980.

14. S. K. Stein, Mathematics, The Man-Made Universe, W. H. Freeman, San Francisco, 1976.

Quantum Mechanics in Hilbert Space, Second Edition. By Eduard Prugovecki. Academic Press, New York, 1981. xxii + 685 pp., $39.50.

JOHN CHALLIFOUR

Department of Mathematics, Indiana University, Bloomington, IN 47405

The power and beauty of mathematical models in physics has often been noted, sometimes with frustration but most often with awe. Quantum mechanics is no exception, and has come to mean those mathematical models or equations which describe the dynamics of the atomic and molecular world. At the mathematical level quantum mechanics describes physical phenomena by assigning to each "idealized state" of a physical system a unit vector in a complex (separable) Hilbert space. Physical measurements, or observables, are related to appropriate self-adjoint operators in Hilbert space, and the actual measured values are directly associated with the spectra of the corresponding operators. Typically, such a unit vector might represent the location of the valence electrons in an atom, and a positive, self-adjoint operator, called the Hamiltonian, has as its spectrum the energy levels which appear as spectral lines in the laboratory.

Since its inception in 1925, nonrelativistic quantum mechanics, independently due to W. Heisenberg and E. Schrödinger, has occupied a central place in the study of self-adjoint operators in Hilbert space by providing nontrivial problems that challenge the usefulness of the general theory and, by their physical setting, offer hints for solution. Among these are the spectral theorem (E. Hellinger, 1909; J. von Neumann, 1929), perturbations of unbounded operators (F. Rellich, 1939; T. Kato, 1957), wave operators and perturbation of absolutely continuous spectra (C. Moller, 1946; K. Friedrichs, 1948; T. Kato and S. T. Kuroda, 1959), and a general theory of operator algebras in Hilbert space (J. von Neumann, 1927; F. J. Murray and J. von

Neumann, 1936–1949). Presence of spin and other discrete quantum numbers in Nature spurred and added greatly to the study of representations of Lie groups (H. Weyl, 1926; E. Wigner, 1939). One might then ask whether after fifty-seven years of intensive and successful research the usefulness of quantum mechanics as a mathematical theory has not in fact been plumbed, leaving possibly difficult but rather specialized questions.

For multiparticle systems (large atoms, electrons in a metal, quantum statistical mechanics) the answer must certainly be no. In fact, quantum statistical mechanics has provided examples of operator algebras in Hilbert space which have been absolutely pivotal in studying and classifying these objects beyond the work of Murray and von Neumann; see, for example, [**1**]. In the two-particle case where the simple Schrödinger equation is at work, there is a close relation between quantum mechanics and diffusion theory in terms of the heat equation; see [**3**], [**5**], [**8**] for precise details. Recent advances in both relativistic and nonrelativistic quantum mechanics have exploited this relation to such an extent that the whole framework of Hilbert space has given way to one of probability theory. Nevertheless there remains a unique feature of quantum mechanics and the Schrödinger equation which as yet has no intrinsic formulation in terms of the related diffusion: the so-called scattering theory describing mathematically the effects of particles scattering off a target. An answer to this question would offer a complete alternative to the Hilbert space formulation of quantum mechanics along probabilistic lines (R. P. Feynman, 1948) and tie yet another branch of mathematics intimately to quantum mechanics.

The last ten years have seen a number of advanced texts on the mathematics and mathematical physics of quantum mechanics, see [**1**], [**3**], [**6**], [**7**], [**8**] among others, and it is clear that this subject has reached a spritely and healthy middle age. However, where could a first-year graduate student in mathematics or physics begin to make his bones in this subject? For one wishing to learn in company with one of the standard physics texts at this level [**4**], formulating the physics rigorously and carrying out many of the standard theoretical physics calculations, Prugovecki's book is a good if somewhat lengthy companion. For such a reader one of this book's strong points is that it is self-contained, starting with vector spaces and measure spaces, taking the reader through the Lebesgue integral and a complete account of linear operator theory on Hilbert space. In fact, this preparation accounts for more than a third of the book, with a middle third on the standard quantum mechanics of Schrödinger and Heisenberg. The main body of application occupies the last third of the text with a complete and extensive treatment of quantum mechanical scattering theory. It is here that most changes appear from the second edition with additions to bring the reader closer to the frontier in this area of research. The reader whose appetite for scattering theory is still not slaked by this account would be ready for volume 3 of [**6**].

The mathematics of quantum mechanics is a heady brew and not for the dilettante, but it is fascinating beyond measure. This book follows closely the syllabus of a first-year graduate course in quantum mechanics in a physics department but with mathematical rigor. For me, it followed that syllabus too closely, and with too much elaboration of similar ground, to set apart those parts of the subject that are solely technical in a mathematical sense from those parts that are conceptual. The one striking omission in a work of this length and focus was the lack of a substantive treatment of the Rayleigh-Schrödinger perturbation theory for the discrete spectrum of the Hamiltonian. After all, this formed the essential physical aspect of quantum mechanics and has prompted much mathematical research; see, for example, the remarks in [**2**]. Nevertheless the dedicated reader can learn much Hilbert space technique and quantum mechanics from this book.

References

1. O. Bratteli and D. W. Robinson, Operator Algebras and Quantum Statistical Mechanics, vol. 1–2, Springer-Verlag, New York, 1979, 1981.

2. W. G. Faris, Bull. Am. Math. Soc., 6 (1982)105–109.

3. J. Glimm and A. Jaffe, Quantum Physics, Springer-Verlag, New York, 1981.

4. A. Messiah, Quantum Mechanics, Wiley, New York, 1962.

5. E. Nelson, Feynman integrals and the Schrödinger equation, J. Math. Phys., 5 (1964)332–343.
6. M. Reed and B. Simon, Methods of Modern Mathematical Physics, vol. 1–4, Academic Press, New York, 1972, 1975, 1978, 1979.
7. M. Schechter, Operator Methods in Quantum Mechanics, Elsevier North-Holland, New York, 1981.
8. B. Simon, Functional Integration and Quantum Physics, Academic Press, New York, 1979.

MISCELLANEA

Russell and Whitehead

96. Symbolism is useful because it makes things difficult. Now in the beginning everything is self-evident, and it is hard to see whether one self-evident proposition follows from another or not. Obviousness is always the enemy to correctness. Hence we must invent a new and difficult symbolism in which nothing is obvious.

Bertrand Russell

It is a safe rule to apply that, when a mathematical or philosophical author writes with a misty profundity, he is talking nonsense.

A. N. Whitehead

—Quoted by R. E. Moritz, *Memorabilia Mathematica*, Macmillan, New York, 1914, pp. 199, 210.

How's that again department

97. The Rasmussen Report estimates there will be one meltdown every 20,000 reactor-years, and one fatality (from cancer) every 50 reactor-years. Conjoin those data (20,000 divided by 50) and you get the figure of 400 deaths a year.

—W. F. Buckley, Chicago Sun-Times, 19 May 1979, p. 36

98. Computation is not barren when it supplies subsistence; it ceases to be an unmitigated evil when it is dignified by ministering to the necessities of material life and comfort. But the computation of the geometer, however tedious it may be, has a loftier aspiration. It provides spiritual nourishment: hence it is life itself, and is the worthy occupation of an immortal soul. The arithmetic formula, considered as an end, is the embodiment of fact, and isolated fact is as worthless as the idle gossip of the parlor or the club; whereas facts combined into formulae and formulae organized into theory penetrate the whole domain of physical science, and ascend to the very throne of ideality.

—Benjamin Peirce, *Ideality in the Physical Sciences*, Little Brown, Boston, 1881.

5. E. Nelson, Feynman integrals and the Schrödinger equation, J. Math. Phys., 5 (1964)332–343.
6. M. Reed and B. Simon, Methods of Modern Mathematical Physics, vol. 1–4, Academic Press, New York, 1972, 1975, 1978, 1979.
7. M. Schechter, Operator Methods in Quantum Mechanics, Elsevier North-Holland, New York, 1981.
8. B. Simon, Functional Integration and Quantum Physics, Academic Press, New York, 1979.

MISCELLANEA

Russell and Whitehead

96. Symbolism is useful because it makes things difficult. Now in the beginning everything is self-evident, and it is hard to see whether one self-evident proposition follows from another or not. Obviousness is always the enemy to correctness. Hence we must invent a new and difficult symbolism in which nothing is obvious.

Bertrand Russell

It is a safe rule to apply that, when a mathematical or philosophical author writes with a misty profundity, he is talking nonsense.

A. N. Whitehead

—Quoted by R. E. Moritz, *Memorabilia Mathematica*, Macmillan, New York, 1914, pp. 199, 210.

How's that again department

97. The Rasmussen Report estimates there will be one meltdown every 20,000 reactor-years, and one fatality (from cancer) every 50 reactor-years. Conjoin those data (20,000 divided by 50) and you get the figure of 400 deaths a year.

—W. F. Buckley, Chicago Sun-Times, 19 May 1979, p. 36

98. Computation is not barren when it supplies subsistence; it ceases to be an unmitigated evil when it is dignified by ministering to the necessities of material life and comfort. But the computation of the geometer, however tedious it may be, has a loftier aspiration. It provides spiritual nourishment: hence it is life itself, and is the worthy occupation of an immortal soul. The arithmetic formula, considered as an end, is the embodiment of fact, and isolated fact is as worthless as the idle gossip of the parlor or the club; whereas facts combined into formulae and formulae organized into theory penetrate the whole domain of physical science, and ascend to the very throne of ideality.

—Benjamin Peirce, *Ideality in the Physical Sciences*, Little Brown, Boston, 1881.

5. E. Nelson, Feynman integrals and the Schrödinger equation, J. Math. Phys., 5 (1964)332–343.
6. M. Reed and B. Simon, Methods of Modern Mathematical Physics, vol. 1–4, Academic Press, New York, 1972, 1975, 1978, 1979.
7. M. Schechter, Operator Methods in Quantum Mechanics, Elsevier North-Holland, New York, 1981.
8. B. Simon, Functional Integration and Quantum Physics, Academic Press, New York, 1979.

MISCELLANEA

Russell and Whitehead

96. Symbolism is useful because it makes things difficult. Now in the beginning everything is self-evident, and it is hard to see whether one self-evident proposition follows from another or not. Obviousness is always the enemy to correctness. Hence we must invent a new and difficult symbolism in which nothing is obvious.

Bertrand Russell

It is a safe rule to apply that, when a mathematical or philosophical author writes with a misty profundity, he is talking nonsense.

A. N. Whitehead

—Quoted by R. E. Moritz, *Memorabilia Mathematica*, Macmillan, New York, 1914, pp. 199, 210.

How's that again department

97. The Rasmussen Report estimates there will be one meltdown every 20,000 reactor-years, and one fatality (from cancer) every 50 reactor-years. Conjoin those data (20,000 divided by 50) and you get the figure of 400 deaths a year.

—W. F. Buckley, Chicago Sun-Times, 19 May 1979, p. 36

98. Computation is not barren when it supplies subsistence; it ceases to be an unmitigated evil when it is dignified by ministering to the necessities of material life and comfort. But the computation of the geometer, however tedious it may be, has a loftier aspiration. It provides spiritual nourishment: hence it is life itself, and is the worthy occupation of an immortal soul. The arithmetic formula, considered as an end, is the embodiment of fact, and isolated fact is as worthless as the idle gossip of the parlor or the club; whereas facts combined into formulae and formulae organized into theory penetrate the whole domain of physical science, and ascend to the very throne of ideality.

—Benjamin Peirce, *Ideality in the Physical Sciences*, Little Brown, Boston, 1881.

FROM THE MAA . . .
A revised edition of a classic

A PRIMER OF REAL FUNCTIONS
Ralph P. Boas, Jr.,
Carus Mathematical Monograph, #13.
Third Edition
xi + 232 pages. Hardbound.
List: $16.50. MAA Member: $12.00

A gold mine of interesting, uncommon insight and examples. . .an orderly composition of 24 partially independent elegant snapshots from the theory of sets and real functions." Lynn A. Steen, commenting on the second edition of "A Primer of Real Functions" in THE AMERICAN MATHEMATICAL MONTHLY, 1974.

The Third Edition includes the most significant revisions to date of this classic volume. Terminology has been modernized, proofs improved, and sections have been completely rewritten. Much new material has been added. The Primer contains the basic material on functions, limits, and continuity that students ought to know before starting a course in real or complex analysis. It is a good place for a student (or anyone else) to see techniques of real analysis at work in uncomplicated but interesting situations.

The author says, "My principal objective was to describe some fascinating phenomena, most of which are not deep or difficult but rarely or never appear in text books or in standard courses." Boas has achieved that objective. The result is a book that should be on every teacher's book shelf, in every college library's, and a part of every serious mathematics student's experience.

Order your copy now from:

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 4 April 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

MATHEMATICAL METHODS OF ECONOMICS*

JOEL FRANKLIN
California Institute of Technology, Pasadena, California 91125

When Dr. Golomb and Dr. Bergquist asked me to give a talk on *economics*, my first impulse was to try to get out of it.

"Sol," I said, "I'm not an economist. You know that."

"I know," said Golomb.

"If you want an economist, I can get you one," I said. "I know some excellent economists."

"No," he said, "we want a mathematician to talk about the subject to other mathematicians from their own point of view."

That made sense, and I hit on this idea: I won't try to tell you what mathematics has done for economics. Instead, I'll do the reverse: I'll tell you some things economics has done for mathematics. I'll describe some mathematical discoveries that were motivated by problems in economics, and I'll suggest to you that some of the new mathematical methods of economics might come into your own teaching and research.

One of these methods is called *linear programming*. I learned about it in 1958. I had just come to Caltech as a junior faculty member associated with the computing center. The director and I made a cross-country trip to survey the most important industrial uses of computers. In New York, we visited the Mobil Oil Company, which had just put in a multi-million-dollar computer system. We found out that Mobil had paid off this huge investment in *two weeks* by doing linear programming.

Back at Caltech, Professor Alan Sweezy in economics and Professors Bill Corcoran and Neil Pings in chemical engineering urged me to teach a course in linear programming. When I told them I didn't know linear programming, they said: Fine, Joel, *learn* it. Seeing they meant business, I did study the subject and give the course. The students loved it, and so did I. Perhaps you will have a similar experience.

One surprising thing I found was this: the mathematics was delightful. I knew it was useful, but I hadn't expected it to be beautiful. I was surprised to find that linear programming wasn't just business mathematics or engineering mathematics; it was the general mathematics of linear inequalities. Later I found this mathematics coming into some of my own special fields of research (statistics, numerical analysis, ill-posed problems). Here again, you may have a similar experience.

Linear programming is one of the many mathematical methods of economics. Here are a few others: quadratic programming, geometric programming, general nonlinear programming; fixed-point theorems—especially the Kakutani theorem; calculus of variations, control theory, dynamics programming; theory of convex sets—especially convex cones; probability, statistics, stochastic processes; finite structures (graph theory, lattice theory); matrix theory; calculus, ordinary differential equations; and special topics like game theory and Arrow's theory of rational preference orderings.

Plato said mathematics is the essence of reality; Willard Gibbs said mathematics is the language of science. If they are right, we shouldn't be surprised to find uses for any branch of

---

The author is Professor of Applied Mathematics at the California Institute of Technology. At Stanford University, in 1953, he received his Ph.D. in pure mathematics. His published research pertains to numerical analysis, to ill-posed problems, to stochastic processes, and to mathematical problems in engineering, in crystallography, in geophysics, and in cell biology. At Caltech he has twice received the Associated Students Award for Excellence in Teaching. He has written two textbooks for undergraduates in mathematics: *Matrix Theory* and *Methods of Mathematical Economics*.

*Invited address to the Mathematical Association of America and to the Society for Industrial and Applied Mathematics, November 14, 1981, Santa Barbara, California.

mathematics in any science. Every branch of mathematics may have some use in the science of economics. Here are two bizarre examples:

Have you heard of *nonstandard analysis*? I've heard of it, but know next to nothing about it. Nevertheless, on November 10, 1981, I heard Yale Economics Professor Donald J. Brown give a colloquium on the *nonstandard analysis of hyper-finite economies* (see [4] and [20]).

You have heard of Bourbaki; so have I. I always thought that stuff would never be good for anything. Nevertheless, Bourbaki *ultrafilters* appear in a paper in the *Journal of Economic Theory* [17]. The authors, A. Kirman and D. Sondermann, use ultrafilters to generalize Kenneth Arrow's fundamental theorem of welfare economics [1].

Mathematics appears in all parts of economics, especially in *mathematical economics* and in *econometrics*. Mathematical economics is like mathematical physics: it is theoretical, nonempirical, sometimes speculative. For instance, Alfred Marshall hypothesized the *existence* of certain curves (supply and demand schedules) whose intersections determine commodity prices. Very pretty, but he didn't show how to measure or predict *numerical* values for specific supply-demand schedules.

In general, *measurement* and *prediction* belong to *econometrics*. As you would expect, econometrics uses a lot of mathematical statistics, probability theory, and numerical analysis. A Nobel prize was given in 1980 to Lawrence Klein for his work in building econometric models.

In 1969 the first Nobel prize in economics was given to Ragnar Frisch and Jan Tinbergen "for having developed and applied dynamic models for the analysis of economic processes"; in other words, the prize was given for mathematics applied to economics. Later, I'll show you a list of all the Nobel prizes in economics, and you'll see that at least 7 of the 12 prizes given from 1969 through 1981 were given for work that could be called applied mathematics. In fact, in 1975 a Nobel prize in economics was given to Leonid Kantorovich, who *is* a mathematician.

In 1969 a spokesman for the Nobel foundation welcomed the new prize subject, economics, as "the oldest of the arts, the youngest of the sciences." It might be fair to say that economics became a science when it started making significant use of mathematics. When was that? I'd say the nineteenth century.

In 1817 the stockbroker David Ricardo proved a *theorem* that establishes an astounding principle of international economics. Ricardo proved mathematically that *free trade is (under certain assumptions) advantageous to consumers in all nations.*

Alfred Marshall was another great nineteenth-century economist. Marshall started out to be a mathematician; he was First Wrangler in mathematics at Cambridge. Although his work is seldom explicitly mathematical, any mathematician reading it can sense its mathematical core. Marshall was a teacher of John Maynard Keynes, whose work contains plenty of explicit mathematics. But, at least to my taste, Marshall's work shows more mathematical insight.

As Gerard Debreu wrote in his *Theory of Value* [7], mathematical economics has become increasingly geometric and qualitative. If we want precise numerical information, we have to turn to *econometrics*. Whereas Marshall drew his supply-and-demand curves in a nonnumerical, qualitative way, the econometrician would have the hard problem of giving *numerical* values for these curves for specific commodities at specific times.

An example of econometrics appears in an article [29] by the mathematician Jacob Schwartz. He used a Wharton econometric model for residential housing. You can see it in Fig. 1. There you see a typical awful equation of econometrics; please don't try to understand it. I just want you to see what it looks like. It predicts the rate of investment in residential housing as a function of various factors (the numerical subscripts refer to time lags). The coefficients (58.26, 0.0249, etc.) come from a numerical curve fit to data for 1948–1964; the model was published in 1967.

There is an old Chinese proverb: *It is always difficult to predict—especially the future*. For that reason econometrics is difficult. The Wharton model of 1967 "predicts" housing starts for 1948–1964—not for the future. In general, econometric models are not laws of nature like $f = ma$ or $E = mc^2$; they are empirical studies whose predictive value depends on the constancy of the underlying relationships.

1967 Wharton econometric model (for 1948–1964)

$$I_h = 58.26 + 0.0249Y - 45.52\left(\frac{p_h}{p_r}\right)_{-3} + 1.433(i_L - i_s)_{-3} + 0.0851\,(I_h^s)_{-1}$$

I_h = rate of investment (\$$10^9$) in residential housing per quarter (3 months)

Y = total disposable income

p_h = average housing price

p_r = average rental price

i_L = long-term interest rate

i_s = short-term interest rate

I_h^s = rate of housing starts

Negative subscripts denote time lags.

FIG. 1.

What Do Economists Think of Mathematics? That question has had different answers at different times. *Now* the answer would be overwhelmingly favorable, if not unanimous. But not so in the old days. Adam Smith published his great book *Wealth of Nations* in 1776. It is readable, fascinating, and important; but it contains almost no mathematics.

I told you the great nineteenth-century economist Alfred Marshall had been First Wrangler in mathematics at Cambridge. Later, he talked about the role mathematics played in his work:

> I had a growing feeling in the later years of my work at the subject that a good mathematical theorem dealing with economic hypotheses was very unlikely to be good economics: and I went more and more on the rules—(1) Use mathematics as a shorthand language, rather than as an engine of inquiry. (2) Keep to them till you have done. (3) Translate into English. (4) Then illustrate by examples that are important in real life. (5) Burn the mathematics. (6) If you can't succeed in 4, burn 3. This last I did often. —quoted in [**31**], p. 307.

So Marshall practiced mathematics as a secret vice; he was a closet mathematician. His most famous student was John Maynard Keynes. At Cambridge, Keynes took his degree in mathematics*. In 1920 Keynes published his *Treatise of Probability*. Keynes's great books on economics contain many equations. By the time of Lord Keynes mathematics was not a secret vice but a public virtue.

A living disciple of Keynes, Harvard Professor John Kenneth Galbraith, regards mathematics with scepticism. One of Galbraith's more entertaining books is called *Economics, Peace, and Laughter*. Commenting on the models of mathematical economics, he says this:

> Moreover, the models so constructed, though of no practical value, serve a useful academic function. The oldest problem in economic education is how to exclude the incompetent.... The requirement that there be an ability to master difficult models, including ones for which mathematical competence is required, is a highly useful screening device.

Not satisfied with this comment, Galbraith adds a dour footnote:

> There can be no question, however, that prolonged commitment to mathematical exercises in economics can be damaging. It leads to the atrophy of judgment and intuition....

John Galbraith does not stand alone. He tells this story about Paul Samuelson, a superb applied mathematician and winner of the Nobel Prize for work in mathematical economics:

*While studying for the Tripos, Keynes wrote to his friend B. W. Swithinbank on 18 April 1905: "I am soddening my brain, destroying my intellect, souring my disposition in a panic-stricken attempt to acquire the rudiments of the Mathematics." See R. F. Harrod [**13**], p. 130.

> Professor Samuelson, in his presidential address to the American Economic Association several years ago, noted that the three previous presidential addresses had been devoted to a denunciation of mathematical economics and that the most trenchant had encouraged the audience to standing applause.

Well! And skepticism about mathematics is not confined to this continent. Galbraith says:

> Once when I was in Russia on a visit to Soviet economists, I spent a long afternoon attending a discussion on the use of mathematical models in plan formation. At the conclusion an elderly scholar, who had also found it very heavy going, asked me rather wistfully if I didn't think there was still a "certain place" for the old-fashioned Marxian formulation of the labor theory of value.

The old Russian scholar must have sighed when a Nobel prize in economics was given to Leonid Kantorovich, a mathematician. Kantorovich got the prize for developing the mathematical theory of linear programming and for applying it to the economic problem of optimum allocation of resources. He would have gone a lot farther with linear programming if he hadn't run into trouble from the orthodox Marxians, who objected to the use of the idea of prices. Dantzig tells the story in his book [**6**], p. 23.

Among the Nobel Laureates in economics, some, like Kantorovich, solved problems in economics by inventing new mathematics; others made much use of known mathematics. Look at the list of Nobel prizes in economics, Fig. 2. I've put asterisks by seven of the twelve prize years to indicate work that is heavily mathematical.

Nobel Prizes in Economics

1969*	Frisch, Ragnar and Tinbergen, Jan—"for having developed and applied dynamic models for the analysis of economic processes."
1970*	Samuelson, Paul—"for the scientific work through which he has developed static and dynamic economic theory and actively contributed to raising the level of analysis in economic science."
1971	Kuznets, Simon—"for his empirically founded interpretation of economic growth which has led to new and deepened insight into the economic and social structure and process of development."
1972*	Hicks, Sir John R. and Arrow, Kenneth J.—"for their pioneering contributions to general economic equilibrium theory and welfare theory."
1973	Leontief, Wassily—"for the development of the input-output method and for its application to important economic problems."
1974	Myrdal, Gunnar and Von Hayek, Friedrich August—"for their pioneering work in the theory of money and economic fluctuations and for their penetrating analysis of the interdependence of economic, social and institutional phenomena."
1975*	Kantorovich, Leonid and Koopmans, Tjalling—"for their contributions to the theory of optimum allocation of resources."
1976*	Friedman, Milton—"for his achievements in the fields of consumption analysis, monetary history and theory and for his demonstration of the complexity of stabilization policy."
1977	Ohlin, Bertil and Meade, James—"for their pathbreaking contributions to the theory of international trade and international capital movements."
1978	Simon, Herbert A.—"for his pioneering research into the decision-making process within economic organizations."
1979	Lewis, Arthur and Shultz, Theodore—for studies of human capital.
1980*	Klein, Lawrence—for computer models designed to forecast economic changes.
1981*	Tobin, James—for mathematical models of investment decisions.

*Asterisks indicate very mathematical work.

FIG. 2.

Seven out of twelve Nobel prizes—not a bad score for mathematics. And some of this mathematics has freshness and charm. For example, let me show you a theorem that won a Nobel prize: the Possibility Theorem of Kenneth Arrow.

In 1957 Kenneth Arrow published a little book called *Social Choice and Individual Values*. He was thinking about a problem of welfare economics: Confronted by numerous conflicting special interests, how should the government make decisions?

Use old-fashioned majority rule, you say. That's the democratic way, isn't it? That's the *rational* way.

Let's see. Suppose we have 3 *alternatives*: vanilla (V), chocolate (C), and strawberry (S). And suppose we have 9 voters, each with his own *individual values*. For example, one individual may like vanilla better than chocolate ($V > C$), and he may like chocolate better than strawberry ($C > S$); then, by the way, he must like vanilla better than strawberry ($V > S$) if his individual values are *rational*. Another individual may prefer strawberry to vanilla ($S > V$), vanilla to chocolate ($V > C$), and *therefore* strawberry to chocolate ($S > C$). And so on.

If all of our nine voters have definite flavor preferences, the voters constitute 6 special-interest groups, corresponding to the six ways of ranking 3 flavors. For example, we might have the following tabulation:

Individual values	Number of individuals
$V > C > S$	2
$S > V > C$	2
$C > S > V$	2
$V > S > C$	1
$C > V > S$	1
$S > C > V$	1

Now comes the general election. Here are the results:

$$V > C \text{ by a majority of 5 to 4}$$
$$C > S \text{ by a majority of 5 to 4}$$

and—what's this?

$$S > V \text{ by a majority of 5 to 4.}$$

But that's crazy: $V > C$ and $C > S$ *should* imply $V > S$, not $S > V$. (This is an example of *Concordet's paradox.*)

No wonder Congress is confused. You see the problem. So did Arrow, and he wondered if there was any way out.

There *is* one way out: Hitler's way. Pick one individual, call him *der Führer*, and do what he says. Then all the government's preferences can be nice and transitive, and too bad for you if you don't like it.

Is there any rational way to make social choices besides dictatorship? To this basic question of welfare economics, Kenneth Arrow gave an astonishing answer: *No.*

ARROW'S THEOREM. *Suppose we have a function that makes rational (transitive) social choices as a function of rational individual values that rank (by preference or indifference) three or more alternatives. Assume that the social-choice function has two properties:*

(i) *If all individuals prefer alternative a to alternative b, then society shall prefer a to b.*

(ii) *The social choice between any two alternatives a and b shall depend only on the individual values between a and b (and should not depend on any third alternative c).*

Then Arrow's theorem says there exists a dictator—a single individual whose preferences become social choices.

In a minute I'll write this theorem symbolically, in terms of matrices. But first I want to explain the two assumptions. The first is a principle of *unanimity*: If everyone prefers vanilla to chocolate, so should society. The second is a principle of *relevance*: Society's choice between vanilla and chocolate should depend on how people feel about vanilla and chocolate, not on how they feel about strawberry.

If you wish, you can write Arrow's theorem in terms of matrices. Let $a_{ij} = 1$ if i is preferred to j; let $a_{ij} = -1$ if j is preferred to i; let $a_{ij} = 0$ if neither is preferred to the other. If there are m alternatives (flavors), then the numbers a_{ij} constitute an $m \times m$ skew-symmetric matrix, A. In a rational preference ordering, if i is preferred to j, and if j is preferred to k, then i must be preferred to k. For the matrix A this says: If $a_{ij} = 1$ and $a_{jk} = 1$, then $a_{ik} = 1$. We shall also require $a_{ik} = 1$ if $a_{ij} = 0$ and $a_{jk} = 1$ *or* if $a_{ij} = 1$ and $a_{jk} = 0$. If this is so, then we'll call A a *rational preference matrix.*

EXAMPLE. Suppose we prefer flavor 3 to flavor 1 and flavor 2, which we like equally. Then this is our rational preference matrix:

$$A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}$$

EXAMPLE. Suppose we prefer flavor 1 to flavor 2, flavor 2 to flavor 3, *and* flavor 3 to flavor 1. That is irrational, and so the preference matrix is *irrational*:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Look: $a_{12} = 1$ and $a_{23} = 1$, *but* $a_{13} \neq 1$.

Individual values and social choice: Suppose there are n individuals and m alternatives. The individual values are expressed by n rational preference matrices $A(1), \ldots, A(n)$. A social choice is a rational preference matrix A. We're looking for a *function* F mapping P_m^n into P_m, where P_m is the set of $m \times m$ rational preference matrices and P_m^n is the n-fold Cartesian product:

$$A = F(A(1), \ldots, A(n)).$$

EXAMPLE. For majority rule, the function F is defined as follows:

$$a_{ij} = \operatorname{sign}\left[a_{ij}(1) + \cdots + a_{ij}(n)\right] \quad (i, j = 1, \ldots, m).$$

If $m > 2$, majority rule may give irrational social choices, as we saw in the example of vanilla, chocolate, and strawberry. So this F takes values outside P_m; but this F does satisfy the assumption of *unanimity* and *relevance*:

$$a_{ij} = 1 \text{ if } a_{ij}(k) = 1 \ \forall k = 1, \ldots, n \tag{1}$$

$$a_{ij} = \text{a function of } a_{ij}(1), \ldots, a_{ij}(n). \tag{2}$$

Arrow's theorem now takes this form: *Let F be a function mapping P_m^n into P_m. Suppose $m > 2$, and suppose the function F satisfies equations* (1) *and* (2). *Then there exists an integer d such that $a_{ij} = 1$ if $a_{ij}(d) = 1$.* (*The integer d depends on F but not on the matrices* $A(1), \ldots, A(n)$.)

By the way, there are no restrictions on the number of individuals, n. In marriage, $n = 2$. Then Arrow's theorem says: Either the husband or the wife must be a dictator, or there must be irrational choices. Experience seems to bear this out.

Arrow's theorem talks about *rational* (transitive) preference orderings. This raises a question in combinatoric analysis: How many rational preferences orderings of m alternatives are there? The answer has appeared in **[12]**. For large m the number of rational preference orderings behaves like $(1/2)m!(\log 2)^{-m-1}$.

The mathematics of Arrow's theorem is very different from mathematics like linear programming. Here we have a rather ordinary looking problem:

For $i = 1, \ldots, m$ and $j = 1, \ldots, n$ we are given the real numbers a_{ij}, b_i, c_j. We wish to find numbers $x_j \geqslant 0$ such that

$$\sum_{j=1}^{n} c_j x_j = \text{minimum}$$

over all solutions of the linear equations

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = 1, \ldots, m).$$

That is the canonical form of linear programming. In terms of matrices and vectors, it looks like this:

$$Ax = b, \quad x \geqslant 0, c^T x = \min.$$

The problem is interesting only if the linear system $Ax = b$ has more than one solution, so we usually suppose rank $A = m < n$. Then the crucial assumption is the sign constraint $x \geqslant 0$ (all components of x must be nonnegative).

Kantorovich in Russia and Dantzig in the United States independently developed linear programming to solve economic logistical problems. The history of their work appears in Dantzig's book **[6]**.

The most famous early problem of linear programming, the *diet problem*, first appeared in the *Journal of Farm Economics* **[33]**. The problem is to design a nutritionally adequate diet at minimum cost. The author, George Stigler, won the 1982 Nobel Prize in Economics.

Suppose a_{ij} is the amount of nutrient i in one unit of food j. (For instance, a_{37} might be the amount of vitamin B_1 in one gram of wheat bread.) Let b_i be the minimum daily requirement of nutrient i, and let c_j be the cost of one unit of food j. Let x_j be the amount of food j in a daily diet. Then we require

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i (i = 1, \ldots, m), \quad x_j \geqslant 0 \quad \sum_{j=1}^{n} c_j x_j = \text{minimum}.$$

This is a linear program in *standard* form. To put it in canonical form, we must replace the m linear inequalities by equations. We do that by introducing m new unknowns $z_i \geqslant 0$:

$$\sum_{j=1}^{n} a_{ij} x_j - z_i = b_i.$$

The problem is now easy to solve by Dantzig's *simplex method.*

Linear programming has many uses in industry and banking. In 1981, a good popular article **[2]** appeared in *Scientific American*; I recommend its example on beer. An introduction to the use of linear programming for the optimization of bank investment portfolios appeared in the *Monthly Review* of the Federal Reserve Bank of Richmond (see **[3]** and **[10]**, p. 3). Banks and oil companies make a lot of money with linear programming.

But you and I are mathematicians; money means nothing to us. So let us speak of something more important—let's talk about *Chebyshev approximation.*

Suppose we are given a system of real linear equations, $Ax = b$, and suppose the system has no solution x. Typically, this occurs when we have more equations than unknowns. If we have m equations in n unknowns, the error in equation i is a function of the vector x:

$$e_i = \sum_{j=1}^{n} a_{ij} x_j - b_i \quad (i = 1, \ldots, m).$$

The problem of Chebyshev approximation is to find a vector x that minimizes the maximum absolute error:

$$\underset{x}{\text{Minimize}}\left(\max_{i}|e_i|\right).$$

That is a beautiful and important problem of approximation theory. Many things were known about Chebyshev approximation before 1959, but no one knew a good way to *do* it. Then Edward Stiefel discovered how to do it by linear programming (see [**32**] and [**10**], p. 8). Here's how:

Define a new unknown: $x_0 = \max|e_i|$ for $i = 1,\dots, m$. Then we shall have the uniform error bracket

$$-x_0 \leqslant \sum_{j=1}^{n} a_{ij}x_j - b_i \leqslant x_0 \quad (i = 1,\dots, m).$$

The problem of Chebyshev is to choose $x_0,\dots, x_n$ so as to minimize the maximum absolute error:

$$\text{Minimize } x_0.$$

That's all there is to it—a finite number of linear inequalities in a finite number of unknowns, with a linear form to be minimized. That is a linear program in general form. It's trivial to restate it in canonical form, and it's routine to solve it numerically by the simplex method.

The simplex method is perhaps the most important numerical method invented in the twentieth century. Experience with enormous industrial problems shows that the simplex method works *fast*. In problems with m equations in n unknowns, the computation time seems to be proportional to n.

Why does the simplex method usually work so fast? No one knows, and this is one of the great unsolved problems of numerical analysis. At first glance, the computation time would seem to be proportional to the binomial coefficient $\binom{n}{m}$, which is the possible number of basic solutions of $Ax = b$. For $m \sim n/2$, the binomial coefficient is almost as big as 2^n, and this suggests the computing time could grow exponentially with n. Indeed, Victor Klee and George Minty [**18**] have constructed pathological cases for which that happens. But it never seems to happen in practice.

A Russian mathematician named Khachian got around this problem by analyzing a quite different algorithm [**16**]. Khachian proved that his algorithm has computing time bounded by a constant, K, times n^6—which becomes smaller than 2^n. Khachian's proof is a triumph of theoretical computer science. But Khachian's algorithm, in its present form, has little practical value: the constant K is enormous and so is the computing time.

You can become famous by doing one of these two things: (1) show why the simplex method usually works as well as it does; (2) show how Khachian's method can be made to work *better* than the simplex method in practice. [A persistent rumor says Stephen Smale *has* done (1).]

Linear programming is important because it is the general mathematics of finite systems of linear inequalities. Linear programming is more general than real linear algebra, for this reason: Any real linear equation $\Sigma a_i x_i = b$ can be restated as a pair of linear inequalities:

$$\sum a_i x_i \leqslant b \quad \text{and} \quad \sum a_i x_i \geqslant b.$$

But the converse is false: You can't restate a linear inequality as a finite number of linear equations.

No mathematician doubts the importance of linear algebra. So linear programming must also be important, and perhaps you will agree that linear programming should be part of the basic undergraduate mathematics curriculum. Why should mathematics students have to pick up their linear programming from economists and chemical engineers and people like that? They should learn it from *us*, and they should learn it right.

Marshall Hall has a section on linear programming in his book *Combinatoric Analysis*. There's nothing odd about that; linear programming has many applications to combinatorics. For

instance, look at this problem:

We are given an $n \times n$ matrix of real numbers a_{ij}. We seek a permutation $j_1, \ldots, j_n$ that maximizes the sum

$$s = a_{1j_1} + a_{2j_2} + \cdots + a_{nj_n}.$$

This problem is called the *optimal-assignment* problem.

EXAMPLE. Suppose we're given the matrix

$$\begin{pmatrix} 7 & 2 & 6 \\ 3 & 9 & 1 \\ 8 & 4 & 5 \end{pmatrix}.$$

The sum s has six possible values. The largest is

$$\max s = a_{13} + a_{22} + a_{31} = 6 + 9 + 8 = 23,$$

achieved for the permutation $(j_1, j_2, j_3) = (3, 2, 1)$.

In general, we could solve the problem by calculating all the $n!$ possible values for s, but that takes too long if n is large. A much faster algorithm is given by linear programming.

We define the unknowns x_{ij} as 1 if $j = j_i$, or 0 if $j \neq j_i$. Thus, x_{ij} will tell us which component to pick from each row. For the preceding numerical example, we would have

$$(x_{ij}) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

In general, the integer unknowns x_{ij} must satisfy the constraints

$$\begin{aligned} &x_{i1} + \cdots + x_{in} = 1 \quad (i = 1, \ldots, n) \\ &x_{1j} + \cdots + x_{nj} = 1 \quad (j = 1, \ldots, n) \\ &x_{ij} \geqslant 0 \quad (i, j = 1, \ldots, n). \end{aligned}$$

Then we wish to maximize a linear form:

$$s = \sum_{i,j} a_{ij} x_{ij} = \text{maximum}.$$

This is a problem in linear programming. H. W. Kuhn **[19]** has shown that it can be solved in $O(n^3)$ steps.

You are right if you object that linear programming provides the optimal *real* solution x_{ij}, and these numbers might not be integers (we need all $x_{ij} = 0$ or 1). But for the optimal-assignment problem the optimal solution over the integers x_{ij} is also optimal over the real numbers x_{ij}. That's not obvious, but it's easy to prove. In general, however, linear programming over the integers is difficult. The optimal solution over integers is usually *not* optimal over real numbers.

So much for combinatorics. Now let's look at geometry. I'd like to show you how *quadratic* programming solves a problem stated in 1857 by J. J. Sylvester **[34]**: "It is required to find the least circle which shall contain a given set of points in the plane."

Suppose the given points are $\mathbf{a}_1, \ldots, \mathbf{a}_n$. We're looking for a circle with the unknown center $\mathbf{x}$ and radius ρ. The given points are required to lie inside the circle:

$$\|\mathbf{a}_i - \mathbf{x}\|^2 \leqslant \rho^2 \quad (i = 1, \ldots, m).$$

Then we want to choose $\mathbf{x}$ and ρ so as to minimize ρ.

We can replace the m quadratic inequalities by *linear* inequalities as follows. Introduce the unknown

$$x_0 = \frac{1}{2}(\rho^2 - \|\mathbf{x}\|^2).$$

Then the m inequalities become

$$x_0 + \mathbf{a}_i \cdot \mathbf{x} \geqslant b_i \quad (i = 1, \ldots, m),$$

where $b_i = \frac{1}{2}\|\mathbf{a}_i\|^2$. Then we want to minimize ρ^2:

$$2x_0 + \|\mathbf{x}\|^2 = \text{minimum}.$$

Sylvester's problem now has this form: First we require m linear inequalities:

$$x_0 + a_{i1}x_1 + a_{i2}x_2 \geqslant b_i \quad (i = 1, \ldots, m).$$

Then we want

$$2x_0 + x_1^2 + x_2^2 = \text{minimum},$$

in which the quadratic terms constitute a positive definite form. This is a routine problem of quadratic programming. It can be solved numerically by an ingenious variant of the simplex method. This algorithm was discovered by a mathematician, Philip Wolfe, but it was published in an *economics* journal, *Econometrica* [**36**].

Why in an economics journal? Because Wolfe's paper extended the work of some economists who were interested in the use of quadratic programming to make optimal investment decisions. Wolfe's mathematical discovery solved a problem in economics.

The theoretical basis of linear and nonlinear programming was published in 1902 by a mathematician named Julius Farkas. He gave a long, cumbersome proof of the following proposition, which you might call *the alternative of linear inequalities* (generalizing the Fredholm alternative of linear equations):

THE FARKAS THEOREM. *Let A be a given $m \times n$ real matrix, and let b be a given vector with m real components. Then one, and only one, of the following alternatives is true*:

(i) *the system $Ax = b$ has a solution $x \geqslant 0$ (all components $\geqslant 0$)*;
(ii) *the system of inequalities $y^TA \geqslant 0$ has a solution y satisfying $y^Tb < 0$.*

Indeed, *both* alternatives can't be true, for then we could deduce

$$0 \leqslant (y^TA)x = y^T(Ax) = y^Tb < 0.$$

That's easy; the hard part is to show that *one* of the alternatives must be true. A modern straightforward proof of the Farkas theorem relies on the separating-plane theorem for convex sets (see, e.g., [**10**], p. 56).

The Farkas alternative has many uses outside mathematical economics. I hope to convince you that every mathematician should know the Farkas theorem and should know how to use it. For example, let me show how to use the Farkas theorem to prove the fundamental theorem of finite Markov processes.

THEOREM (Markov). *Suppose $p_{ij} \geqslant 0$, and suppose*

$$\sum_{i=1}^{n} p_{ij} = 1 \quad (j = 1, \ldots, n).$$

Then there exist numbers $x_j \geqslant 0$ satisfying

$$\sum_{j=1}^{n} p_{ij}x_j = x_i \quad (i = 1, \ldots, n)$$

$$\sum_{j=1}^{n} x_j = 1.$$

The proof of a special case of this theorem occupies several pages in Feller's book on probability ([**8**], pp. 428–432). The general case is usually proved by using the Perron-Frobenius maximum principle for positive matrices or by using the Brouwer fixed-point theorem. Instead, we can give an elementary proof using the Farkas theorem ([**10**], p. 58):

First, we state Markov's assertion as one Farkas alternative:

(i) There exists a vector $x \geqslant 0$ satisfying the $n + 1$ linear equations

$$\begin{aligned} \sum_{j=1}^{n} (p_{ij} - \delta_{ij})x_j &= 0 \quad (i = 1,\ldots, n) \\ \sum_{j=1}^{n} x_j &= 1, \end{aligned}$$

where δ_{ij} is the Kronecker delta.

Second, we state the other Farkas alternative:

(ii) There exist numbers $y_1,\ldots, y_n, y_{n+1}$ satisfying the inequalities

$$\begin{aligned} \sum_{i=1}^{n} y_i\left(p_{ij} - \delta_{ij}\right) + y_{n+1} &\geqslant 0 \quad (j = 1,\ldots, n) \\ y_{n+1} &< 0. \end{aligned}$$

Alternative (ii) implies the strict inequalities

$$\sum_{i=1}^{n} y_i p_{ij} > y_j \quad \text{for all } j.$$

But

$$\max y_i \geqslant \sum_{i=1}^{n} y_i p_{ij}$$

because we assumed $p_{ij} \geqslant 0$ and $\Sigma_i\, p_{ij} = 1$, so we find

$$\max y_i > y_j \quad \text{for all } j.$$

That is impossible, so alternative (ii) is false.

Now Farkas tells us that alternative (i) is true: Markov's theorem is proved. That was easy, wasn't it?

Now let me tell you about *the theory of games and economic behavior*. A book with that title was published in 1944 by the mathematician John von Neumann and the economist Oskar Morgenstern [**25**]. Economists consider this book an epoch-making contribution to economics.

Fine, you say, but *what has it done for mathematics*?

This book, along with von Neumann's earlier work [**24**] on game theory, has given us some stimulating problems and some important results. For example, look at this theorem on matrices:

THEOREM (VON NEUMANN). *Let A be a real $m \times n$ matrix. Let vectors x and y range over the sets*

$$\sum_{i=1}^{m} x_i = 1, \quad x_i \geqslant 0; \quad \sum_{j=1}^{n} y_j = 1, \quad y_j \geqslant 0.$$

Then

$$\min_y \max_x x^T A y = \max_x \min_y x^T A y.$$

This theorem is no platitude. As a rule, mixed extrema are *not* equal, as the following example shows. Suppose x and y range over the sets $0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1$. Then

$$\min_y \max_x (x-y)^2 = \frac{1}{4},$$

but

$$\max_x \min_y (x-y)^2 = 0.$$

Von Neumann's *minimax theorem* is the fundamental result in the theory of zero-sum two-person games. But that's not the point; the point is, it's good *mathematics*. Von Neumann proved the minimax theorem by using the Brouwer fixed-point theorem. His proof is nonelementary and nonconstructive. Later, the mathematician George Dantzig gave an elementary, constructive proof by using the dual simplex method of linear programming.

Following von Neumann, mathematical economists make much use of the fixed-point theorems. Their favorite seems to be the fixed-point theorem of Kakutani [**15**].

As a young mathematician at the Institute of Advanced Study, Shizuo Kakutani discovered a generalization of the Brouwer fixed-point theorem. Kakutani's work was motivated by problems in economic game theory. His theorem has great mathematical novelty. It speaks of point-to-set mappings:

THEOREM (Kakutani). *Let X be a closed, bounded, convex set in $\mathbb{R}^n$. For every point $\mathbf{x}$ in X, let $F(\mathbf{x})$ equal a nonempty convex subset of X. Assume that the graph*

$$\{\mathbf{x}, \mathbf{y} : \mathbf{y} \in F(\mathbf{x})\} \quad \text{is closed.}$$

Then some point in X satisfies $\mathbf{x}^* \in F(\mathbf{x}^*)$.

The image of each point $\mathbf{x}$ is a convex set $F(\mathbf{x}) \subset X$. The theorem says some point $\mathbf{x}^*$ lies *in* its image $F(\mathbf{x}^*)$. Figure 3 illustrates this. Kakutani's theorem is novel because it talks about *set*-valued functions.

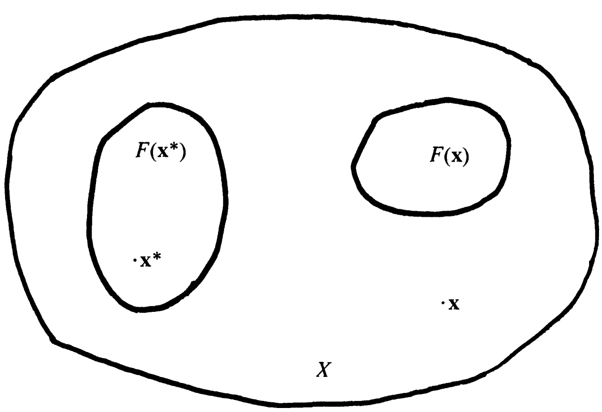


FIG. 3.

If every set $F(\mathbf{x})$ contains just one point, the closed-graph assumption is equivalent to the continuity of the function $F(\mathbf{x})$, and then Kakutani's theorem reduces to the Brouwer fixed-point theorem. Kakutani proved his theorem by using the Brouwer theorem.

A private survey indicates that 96% of all mathematicians can state the Brouwer fixed-point theorem, but only 5% can prove it. Among mathematical economists, 95% can state it, but only 2% can prove it (and these are all ex-topologists). This dangerous situation will soon be remedied. Within the last two years, John Milnor [**22**] and C. A. Rogers [**27**] have produced elementary proofs, using nothing more advanced than calculus. These proofs are so easy that *I* can

understand them [**10**], and certainly *you* can.

While 96% of mathematicians can *state* the Brouwer fixed-point theorem, only 7% can state the Kakutani theorem. This situation is also dangerous or, at least, wasteful. The Kakutani theorem has many potential applications outside economics; these applications should be made. Now that we can all understand the Brouwer theorem, we can also understand the Kakutani theorem, so nothing can stop us.

In the application of Kakutani's theorem to many-person game theory, the point $\mathbf{x}$ denotes a collection of mixed strategies and the set-valued function denotes the sets of optimal mixed strategies. The inclusion $\mathbf{x} \in F(\mathbf{x})$ characterizes an equilibrium solution of the game. The Kakutani theorem is thus the perfect tool for proving J. F. Nash's fundamental theorem [**23**] on n-person games.

Professor H. F. Bohnenblust once told me something about research. He had supervised many successful Ph.D. thesis projects—and a few unsuccessful ones. He said this: The unsuccessful projects start with some famous old problem (prove the Riemann hypothesis) and *then* look for a method to solve it. The successful projects *start with some new method and then look for a problem.*

Let's take Bohnenblust's advice. Let's start with linear programming and look for a problem. Here's a good one: the *problem of moments* in probability theory.

Suppose we are given a collection of real-valued continuous functions $a_i(\mathbf{t})$ for $\mathbf{t} \in \mathbb{R}^p$. We are given a closed set $\Omega \subset \mathbb{R}^p$, and we're given a collection of real numbers b_i. The problem is to find a probability distribution function $x(t)$ satisfying the moment equations

$$\int_\Omega a_i(\mathbf{t})\, dx(\mathbf{t}) = b_i \quad \text{for all } i$$

where we require $dx(\mathbf{t}) \geqslant 0$ and

$$\int_\Omega dx(\mathbf{t}) = 1.$$

This problem has many applications in geophysics and in other sciences. It has an extensive mathematical theory (see, for instance, Shohat and Tamarkin [**30**]). So what is left for you and me to do here? Well, for one thing, we could devise a good numerical method. At least, that will please our colleagues in geophysics.

Suppose we're given a *finite* number of moments, which is the usual case in applications. And suppose we use some numerical scheme to approximate the integrals by finite sums. Then we get a finite set of linear equations in a finite set of unknowns:

$$\sum_{j=1}^{n} a_{ij} x_j = b_j \quad (i = 1, \ldots, m).$$

Now we're looking for the numbers $x_1, \ldots, x_n$; they will constitute a finite set of probabilities, satisfying

$$\sum_{j=1}^{n} x_j = 1, \quad x_j \geqslant 0.$$

So we want to solve $m + 1$ linear equations in n unknowns $x_j \geqslant 0$. Ah! We recognize a problem in linear programming. For this we have an existence theorem, the Farkas theorem, and a numerical method, the simplex method.

The simplex method will tell us if no solution exists, or it will compute a solution x if solutions do exist. For $n > m + 1$ we can't expect the solution x to be unique. We are free to impose any minimum condition of the form

$$\sum_{j=1}^{n} c_j x_j = \text{minimum}.$$

We note that the original problem with a *finite* number of moments usually doesn't have a unique solution $x(\mathbf{t})$, so the freedom to impose an extra condition is physically natural and mathematically necessary.

Fine, you say. All right for some people but not for you. You are a pure mathematician, and numerical methods bore you. What you'd like is a little solid theory—something you can get your teeth into.

OK, I'm with you. Let's prove a great theorem together. Let's give a new, elementary proof of a famous theorem of F. Hausdorff [**14**]. The proof will use a method of mathematical economics, the Farkas theorem.

Hausdorff studied the moment problem

$$\int_0^1 t^k \, dx(t) = b_k \quad (k = 0, 1, \ldots). \tag{3}$$

He asked this question: *Which infinite sequences* $\{b_k\}$ *are the moments of a probability distribution* $x(t)$ *on the interval* $0 \leqslant t \leqslant 1$? He called those sequences *moment sequences.*

Certainly $b_0 = 1$, since we require $\int dx(t) = 1$. Also, we must have

$$\int_0^1 f(t) \, dx(t) \geqslant 0$$

for all continuous functions $f(t) \geqslant 0$. Setting $f(t) = t^j(1 - t)^k$, we get the necessary condition

$$\int_0^1 \sum_{\nu=0}^{k} (-)^\nu \binom{k}{\nu} t^{j+\nu} \, dx(t) \geqslant 0,$$

which says this about the moments:

$$\sum_{\nu=0}^{k} (-)^\nu \binom{k}{\nu} b_{j+\nu} \geqslant 0 \quad (j, k \geqslant 0).$$

A sequence $\{b_i\}$ with this property is called *completely monotone*. If we define the difference operator Δ by $\Delta b_i = b_{i+1} - b_i$, the last formula says

$$(-)^k \Delta^k b_j \geqslant 0 \quad (j, k \geqslant 0).$$

Hausdorff's theorem says: *If* $b_0 = 1$, *the sequence* $b_0, b_1, b_2, \ldots$ *is a moment sequence if and only if it is completely monotone.*

We've already proved the *only if* part. To prove the *if* part, let's assume the sequence $\{b_i\}$ is completely monotone, with $b_0 = 1$. Now we must find a p.d.f. (probability distribution function) $x(t)$ satisfying the moment equations (3).

Suppose we can solve the system of moment equations

$$\int_0^1 t^k \, dx_n(t) = b_k \quad (k = 0, \ldots, n) \tag{i}$$

for each *finite* n. Then the p.d.f.'s $x_n(t)$ have a subsequence that converges to a p.d.f. $x(t)$ at all points of continuity of the limit $x(t)$. Then $x(t)$ satisfies *all* the moment equations (3), and we're done.

So the required p.d.f. $x(t)$ exists unless some *finite* system (i) is unsolvable. But the system (i) is a finite linear system for an unknown $dx_n(t) \geqslant 0$. A simple extension of the Farkas theorem says this: *The system* (i) *is unsolvable for a* p.d.f. $x_n(t)$ *if and only if there exist numbers* $y_0, \ldots, y_n$ *satisfying*

$$\text{(ii)} \qquad \begin{aligned} &\sum_{k=0}^{n} y_k t^k \geqslant 0 \quad (0 \leqslant t \leqslant 1) \\ &\sum_{k=0}^{n} y_k b_k < 0. \end{aligned}$$

We must show that this is impossible.

Suppose (ii) is true. Define the polynomial $f(t) = \Sigma\, y_k t^k$. Then Taylor's theorem says

$$y_k = f^{(k)}(0)/k!$$

As a limit of difference quotients, this equals

$$y_k = \lim_{\varepsilon\to 0} \Delta_\varepsilon^k f(0)/(\varepsilon^k k!),$$

where $\Delta_\varepsilon f(t) = f(t+\varepsilon) - f(t)$. Setting $\varepsilon = 1/N$, we deduce

$$y_k = \lim_{N\to\infty} \binom{N}{k} \Delta_\varepsilon^k f(0).$$

The second part of (ii) says $\Sigma\, y_k b_k < 0$, and so for large N we must have

$$\sum_{k=0}^{n} \binom{N}{k} \left(\Delta_\varepsilon^k f(0)\right) \cdot b_k < 0.$$

The upper limit, n, may be replaced by a larger integer, N, since an nth degree polynomial $f(t)$ satisfies $\Delta_\varepsilon^k f(t) = 0$ for $k > n$. Now we rearrange the last sum to obtain the inequality

$$\sum_{j=0}^{N} f\left(\frac{j}{N}\right)\binom{N}{j} \cdot (-)^{N-j} \Delta^{N-j} b_j < 0.$$

But (ii) says $f \geqslant 0$, and the completely monotone sequence $\{b_i\}$ satisfies $(-)^k \Delta^k b_j \geqslant 0$, so all terms in the last sum are *nonnegative*, and we have a contradiction. The Farkas alternative (ii) is impossible.

Therefore, the alternative (i) is true: every finite system of moment equations (i) is solvable. It follows that the infinite system (3) is solvable, and so we have proved Hausdorff's theorem.

This theorem is important in probability theory. As William Feller said, "Its discovery has been justly celebrated as a deep and powerful result." (See [**9**], p. 226.)

As you've just seen, the mathematical methods of economics have striking applications to the rest of mathematics. As you might have feared, I could go on talking to you forever. I could tell you about applications to ill-posed boundary-value problems of partial differential equations. But I manfully refrain; you have already heard enough. By now, I hope you will agree with me: these problems and methods of economics are valuable, and they are fascinating.

References

1. Kenneth Arrow, Social Choice and Individual Values, 2nd ed., Yale University Press, New Haven, 1963.

2. Robert G. Bland, The allocation of resources by linear programming, Sci. Amer., 244 (1981) 126–144.

3. Alfred Broaddus, Linear programming: a new approach to bank portfolio management, Monthly Review, Federal Reserve Bank of Richmond, Virginia, vol. 58 (1972) 3–11.

4. Donald J. Brown, Nonstandard economies: a survey, to appear in Studies in Mathematical Economics, S. Reiter, editor, Math. Assoc. of America.

5. George B. Dantzig, Constructive proof of the min-max theorem, Pacific J. Math., 6 (1956) 25–33.

6. ______, Linear Programming and Extensions, Princeton Univ. Press, 1963.

7. Gerard Debreu, Theory of Value, Yale University Press, New Haven, 1959.

8. William Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., Wiley, New York, 1968.

9. ______, An Introduction to Probability Theory and Its Applications, vol. 2, 2nd ed., Wiley, New York, 1971.

10. Joel Franklin, Methods of Mathematical Economics, Springer-Verlag, New York, 1980.

11. John Kenneth Galbraith, Economics, Peace, and Laughter, essays edited by Andrea D. Williams, Houghton Mifflin Company, New York, 1971.

12. O. A. Gross, Preferential arrangements, this MONTHLY, 69 (1962) 4–8.

13. R. F. Harrod, The Life of John Maynard Keynes, Avon Books, New York, 1951.

14. F. Hausdorff, Momentprobleme für ein endliches Intervall, Math. Z., 16 (1923) 220–248.

15. S. Kakutani, A generalization of Brouwer's fixed-point theorem, Duke Math. J., 8 (1941) 457–458.

16. L. V. Khachian, A polynomial algorithm in linear programming, Soviet Math. Dokl., 20 (1979) 191–194 (Amer. Math. Soc. translation).

17. Alan P. Kirman and Dieter Sondermann, Arrow's theorem, many agents, and invisible dictators, J. Econom. Theory, 5 (1972) 267–277.

18. Victor Klee and George Minty, How good is the simplex algorithm? Inequalities, 3 (1972) 159–175, Academic Press, New York.

19. H. W. Kuhn, The Hungarian method for the assignment problem, Naval Res. Logistics Quart., 2 (1955) 83–97.

20. W. A. J. Luxemburg and K. Stroyan, The Theory of Infinitesimals, Academic Press, New York, 1976.

21. G. S. Maddala, Econometrics, McGraw-Hill, New York, 1978.

22. John Milnor, Analytic proofs of the "hairy ball theorem" and the Brouwer fixed-point theorem, this MONTHLY, 85 (1978) 521–524.

23. J. F. Nash, Equilibrium points in n-person games, Proc. Nat. Academy of Sciences, 36 (1950) 48–49.

24. John von Neumann, Zur Theorie der Gesellschaftsspiele, Math. Ann., 100 (1928) 295–320.

25. John von Neumann and Oskar Morgenstern, Theory of Games and Economic Behavior, Princeton Univ. Press, 1944.

26. H. Nikaido, Convex Structures and Economic Theory, Academic Press, New York, 1968.

27. C. A. Rogers, A less strange version of Milnor's proof of Brouwer's fixed-point theorem, this MONTHLY, 87 (1980) 525–527.

28. Paul Samuelson, Economics, 9th ed., McGraw-Hill, New York, 1973.

29. Jacob Schwartz, Mathematics as a tool for economic understanding, pp. 269–295 in Mathematics Today: Twelve Informal Essays, edited by L. A. Steen, Springer-Verlag, New York, 1978.

30. J. A. Shohat and J. D. Tamarkin, The Problem of Moments, Mathematical Surveys, no. 1, American Mathematical Society, New York, 1943.

31. G. F. Shove, The place of Marshall's Principles in the development of economic theory, Econom. J., 52 (1942) 294–329.

32. Eduard Stiefel, An Introduction to Numerical Mathematics, Academic Press, New York, 1963.

33. George Stigler, The cost of subsistence, J. Farm Econom., 27 (1945) 303–314.

34. J. J. Sylvester, A question in the geometry of situation, Quart. J. Pure and Appl. Math., 1 (1857) 79.

35. A. Takayama, Mathematical Economics, Holt, Rinehart and Winston, 1974.

36. Philip Wolfe, The simplex method for quadratic programming, Econometrica, 27 (1959) 382–398.

MISCELLANEA

99. Bolzano-Weierstrass for ornithologists.

We came... to a heavy thicket of bramble and stones, rising like a dinosaur out of which was an extraordinary contrivance of wire fences, wood, and ropes, of doors and pulleys moaning in the breeze. This, Miss Whittaker said cheerfully, was a Heligoland bird trap, the most serviceable kind there is. I stared astonishedly at the Heligoland bird trap, estimating it one hundred feet long, thirty feet wide, and eight feet high—roughly one million times as large as the average bird. Presently, Miss Whittaker saw my dismay, and observed that a Heligoland bird trap... is superior to any [other device] because of the sheer intricacy of its mechanism, which only the most erudite of birds could hope to grasp. Its proper utilization begins, she continued, when she thrashes about in the brambles and frightens a bird into the trap, shutting a screen door behind it. The bird flies about in consternation, and more and more screen doors are closed by Miss Whittaker, who—did I say?—is also inside the Heligoland bird trap; the woebegone bird finds itself in smaller and smaller quarters, and finally, when the last door closes, in a small, accessible, wooden box.

—John Sack, *Report from Practically Nowhere*, Harper, New York, 1959, p. 23. Suggested by H. P. Boas.

15. S. Kakutani, A generalization of Brouwer's fixed-point theorem, Duke Math. J., 8 (1941) 457–458.

16. L. V. Khachian, A polynomial algorithm in linear programming, Soviet Math. Dokl., 20 (1979) 191–194 (Amer. Math. Soc. translation).

17. Alan P. Kirman and Dieter Sondermann, Arrow's theorem, many agents, and invisible dictators, J. Econom. Theory, 5 (1972) 267–277.

18. Victor Klee and George Minty, How good is the simplex algorithm? Inequalities, 3 (1972) 159–175, Academic Press, New York.

19. H. W. Kuhn, The Hungarian method for the assignment problem, Naval Res. Logistics Quart., 2 (1955) 83–97.

20. W. A. J. Luxemburg and K. Stroyan, The Theory of Infinitesimals, Academic Press, New York, 1976.

21. G. S. Maddala, Econometrics, McGraw-Hill, New York, 1978.

22. John Milnor, Analytic proofs of the "hairy ball theorem" and the Brouwer fixed-point theorem, this MONTHLY, 85 (1978) 521–524.

23. J. F. Nash, Equilibrium points in n-person games, Proc. Nat. Academy of Sciences, 36 (1950) 48–49.

24. John von Neumann, Zur Theorie der Gesellschaftsspiele, Math. Ann., 100 (1928) 295–320.

25. John von Neumann and Oskar Morgenstern, Theory of Games and Economic Behavior, Princeton Univ. Press, 1944.

26. H. Nikaido, Convex Structures and Economic Theory, Academic Press, New York, 1968.

27. C. A. Rogers, A less strange version of Milnor's proof of Brouwer's fixed-point theorem, this MONTHLY, 87 (1980) 525–527.

28. Paul Samuelson, Economics, 9th ed., McGraw-Hill, New York, 1973.

29. Jacob Schwartz, Mathematics as a tool for economic understanding, pp. 269–295 in Mathematics Today: Twelve Informal Essays, edited by L. A. Steen, Springer-Verlag, New York, 1978.

30. J. A. Shohat and J. D. Tamarkin, The Problem of Moments, Mathematical Surveys, no. 1, American Mathematical Society, New York, 1943.

31. G. F. Shove, The place of Marshall's Principles in the development of economic theory, Econom. J., 52 (1942) 294–329.

32. Eduard Stiefel, An Introduction to Numerical Mathematics, Academic Press, New York, 1963.

33. George Stigler, The cost of subsistence, J. Farm Econom., 27 (1945) 303–314.

34. J. J. Sylvester, A question in the geometry of situation, Quart. J. Pure and Appl. Math., 1 (1857) 79.

35. A. Takayama, Mathematical Economics, Holt, Rinehart and Winston, 1974.

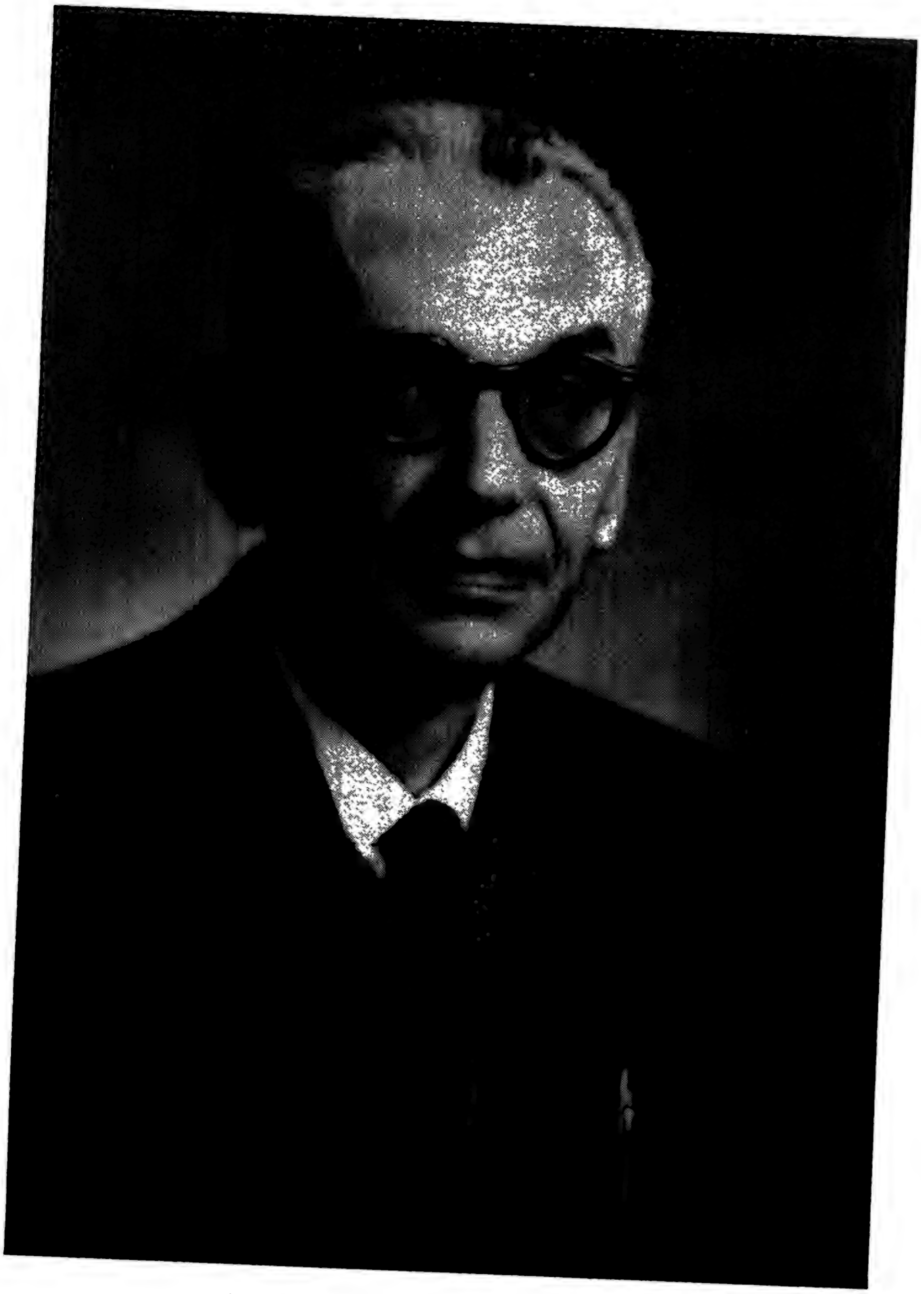
36. Philip Wolfe, The simplex method for quadratic programming, Econometrica, 27 (1959) 382–398.

MISCELLANEA

99. Bolzano-Weierstrass for ornithologists.

We came... to a heavy thicket of bramble and stones, rising like a dinosaur out of which was an extraordinary contrivance of wire fences, wood, and ropes, of doors and pulleys moaning in the breeze. This, Miss Whittaker said cheerfully, was a Heligoland bird trap, the most serviceable kind there is. I stared astonishedly at the Heligoland bird trap, estimating it one hundred feet long, thirty feet wide, and eight feet high—roughly one million times as large as the average bird. Presently, Miss Whittaker saw my dismay, and observed that a Heligoland bird trap... is superior to any [other device] because of the sheer intricacy of its mechanism, which only the most erudite of birds could hope to grasp. Its proper utilization begins, she continued, when she thrashes about in the brambles and frightens a bird into the trap, shutting a screen door behind it. The bird flies about in consternation, and more and more screen doors are closed by Miss Whittaker, who—did I say?—is also inside the Heligoland bird trap; the woebegone bird finds itself in smaller and smaller quarters, and finally, when the last door closes, in a small, accessible, wooden box.

—John Sack, *Report from Practically Nowhere*, Harper, New York, 1959, p. 23. Suggested by H. P. Boas.



To decide this problem, see p. 266.

The numbers in parentheses are the rankings by the Associated Press and the United Press International opinion polls. The first three columns present the win-loss-tie record. The ranking itself is presented in the next column. The numbers represent the ranking on a logarithmic scale base 2 (the scale is calibrated so as the team ranked number 60 has ranking value 0). The p's and x's can be computed from this scale. For instance

$$x_{1,2} = 2^{8.833-8.784} = 2^{0.049} = 1.034$$

and

$$p_{1,2} = \frac{x_{1,2}}{1 + x_{1,2}} = \frac{1.034}{2.034} = .508$$

For each team in the table we also give four selected "odds" x_{ij}. For the nth team we list the numbers

$$x_{n,1},\, x_{n,20},\, x_{n,50} \quad \text{and} \quad x_{n,100}$$

comparing the team with the teams ranked number 1, 20, 50 and 100.

Two technical remarks:

1. All the 177 teams are comparable. One team had a perfect losing record, all other teams form one equivalence class.

2. Some teams played against opponents from Divisions II or III. For the purpose of this ranking, these lower division opponents are identified and treated as one single team. This approximation introduces a slight error into the ranking of the bottom teams; it has, however, no discernible effect on the high ranked teams. [This hypothetical "average Division II/III opponent" is ranked number 155 (with ranking value -3.685). The highest ranked team that played a lower division opponent is number 46, and we have $p(46, 155) = .985$, $x(46, 155) = 67$.]

Appendix 2 The Split 1981 Baseball Season

Due to a strike in the middle of the season it was decided that the 1981 season would be divided into two halves: the games played before the strike and the games played after the strike. The teams with the best winning percentage in each division were declared first-division winners. At the time of this writing, the second half is still being played.

In the table on p. 265 we give the ranking of both leagues, based on their results in the first half of the 1981 season. In each league, we list the Eastern division teams first, in decreasing rank, followed by the Western division teams. The numbers displayed next to each team represent

(1) the schedule, i.e., the number of games played against other teams in the league;
(2) the won-lost record;
(3) the ranking; the numbers p_{ij} rounded off to three decimal places;
(4) the expected score: this is calculated from the p_{ij}, and is based on the complete schedules of both leagues as displayed in (2.3).

In three cases the highest-ranked team in its division is the declared winner. Note, however, that in American League West, the declared winner (the team with the highest win-loss ratio) is ranked second by our method.

ANSWER TO PHOTO ON PAGE 245

Kurt Gödel.

THE RANKING OF INCOMPLETE TOURNAMENTS: A MATHEMATICIAN'S GUIDE TO POPULAR SPORTS

THOMAS JECH
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

1. Introduction. Each winter millions of Americans engage in passionate discussions about which collegiate football team should be the national champion. A handful is even privileged enough to have their opinions tabulated by two opinion polls. These polls then decide who the champion is.

It has been a long tradition in sports that champions are determined, whenever possible, by objective methods, preferably measuring performance numerically, in terms of seconds, meters, number of wins and losses, etc. In team sports, for example, the standard system is a round-robin tournament, where each team plays against every other team (possibly more than once but the same number of times), and counting a win as one point and a tie as a half point, the team with the largest number of points is the winner.

So it is surprising that in a popular sport like collegiate football, the champion is determined by such arbitrary methods. And the obvious question is: *Is there an objective method that ranks the teams according to their records even when their playing schedules are not identical?*

It is the purpose of this article to show that this problem can be formulated mathematically and to give a solution of this problem. We show that (subject to a very natural condition) there is one and only one correct way of comparing the records of teams in an incomplete tournament. The condition under which our method is applicable and which will be made precise in Section 2 can be expressed roughly as: "there are enough results available to have a basis for comparison."

The method itself is not limited to applications in sports. As another example, consider a situation when a researcher considers a certain large number of objects and wants to rank them according to some of their properties but can only make a limited number of comparisons. The test results may even be inconsistent; for instance suggesting that A is better than B, B is better than C and C is better than A.

Moreover, a result of comparing object A with object B may not be conclusive but only give probability p, $0 \leqslant p \leqslant 1$, that A is better (bigger, brighter, more desirable, more marketable) than B. So our researcher has obtained these results r_{ij} ($0 \leqslant r_{ij} \leqslant 1$, $r_{ji} = 1 - r_{ij}$) for some i's and j's and the question is: Do these results provide a ranking of the given items, and in particular, can we tell which one is best (biggest, brightest, most desirable, most marketable) or at least which is most likely to be such? It should be clear that this is, mathematically, the same problem as the problem of ranking college football teams. So our method can also be applied in such situations.

In the appendix I give some applications of my method. One example I use is the 1981 major league baseball season, in which (due to a strike) the teams played an unequal number of games against differing opponents.

2. Formulation of the Problem and Statement of the Result.

Tournaments. Although the ranking method can in general be applied in other situations, such

---

I was born in Prague in 1944. Mathematics has fascinated me since a very early age, and although I have had many other interests over the years, from chess to rock climbing to marathon running, mathematics has always remained my main interest. While in high school I competed in Mathematical Olympiads. At Charles University I became interested in set theory, mostly under the influence of Petr Vopěnka, and in 1966 I earned a Doctorate for work using the then new method of forcing invented by Paul Cohen. I fled Czechoslovakia after the Soviet occupation in 1968 and spent the next six years wandering about British and U.S. universities. Since 1974 I have been a Professor of Mathematics at The Pennsylvania State University. T. J.

I wish to express my gratitude to my wife and son for their invaluable advice on the use of the computer and on baseball and football, of which I am rather ignorant.

as I touched upon in Section 1, I prefer to state it in terms of comparing teams in a competition. Thus let us consider n teams

$$T_1, T_2, \ldots, T_n$$

which are matched in a certain way so that some teams play each other (possibly more than once) and some do not. The information which pairs of teams have played is given by the *schedule matrix*

$$M = (m_{ij})_{i,j=1}^{n}. \tag{2.1}$$

M is a symmetric n by n matrix and each m_{ij} is a nonnegative integer. The number m_{ij} represents the number of times the team T_i is pitted against the team T_j. For each i and j we have

$$m_{ij} = m_{ji}$$

and

$$m_{ii} = 0.$$

As examples, consider a *simple round-robin tournament* whose schedule matrix is

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 0 \end{pmatrix} \tag{2.2}$$

and the schedule matrices of a complete season of American and National League baseball:
American League

$$M_A = \begin{pmatrix}
0 & 13 & 13 & 13 & 13 & 13 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 0 & 13 & 13 & 13 & 13 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 13 & 0 & 13 & 13 & 13 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 13 & 13 & 0 & 13 & 13 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 13 & 13 & 13 & 0 & 13 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 13 & 13 & 13 & 13 & 0 & 13 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
13 & 13 & 13 & 13 & 13 & 13 & 0 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 0 & 13 & 13 & 13 & 13 & 13 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 0 & 13 & 13 & 13 & 13 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 13 & 0 & 13 & 13 & 13 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 13 & 13 & 0 & 13 & 13 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 13 & 13 & 13 & 0 & 13 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 13 & 13 & 13 & 13 & 0 & 13 \\
12 & 12 & 12 & 12 & 12 & 12 & 12 & 13 & 13 & 13 & 13 & 13 & 13 & 0
\end{pmatrix} \tag{2.3}$$

National League

$$M_N = \begin{pmatrix}
0 & 18 & 18 & 18 & 18 & 18 & 12 & 12 & 12 & 12 & 12 & 12 \\
18 & 0 & 18 & 18 & 18 & 18 & 12 & 12 & 12 & 12 & 12 & 12 \\
18 & 18 & 0 & 18 & 18 & 18 & 12 & 12 & 12 & 12 & 12 & 12 \\
18 & 18 & 18 & 0 & 18 & 18 & 12 & 12 & 12 & 12 & 12 & 12 \\
18 & 18 & 18 & 18 & 0 & 18 & 12 & 12 & 12 & 12 & 12 & 12 \\
18 & 18 & 18 & 18 & 18 & 0 & 12 & 12 & 12 & 12 & 12 & 12 \\
12 & 12 & 12 & 12 & 12 & 12 & 0 & 18 & 18 & 18 & 18 & 18 \\
12 & 12 & 12 & 12 & 12 & 12 & 18 & 0 & 18 & 18 & 18 & 18 \\
12 & 12 & 12 & 12 & 12 & 12 & 18 & 18 & 0 & 18 & 18 & 18 \\
12 & 12 & 12 & 12 & 12 & 12 & 18 & 18 & 18 & 0 & 18 & 18 \\
12 & 12 & 12 & 12 & 12 & 12 & 18 & 18 & 18 & 18 & 0 & 18 \\
12 & 12 & 12 & 12 & 12 & 12 & 18 & 18 & 18 & 18 & 18 & 0
\end{pmatrix}$$

The top seven rows of M_A (the top six rows of M_N) belong to the eastern teams and the bottom seven (six) rows to the western teams.

Let T_i and T_j be two teams playing in the tournament. The *result* of the m_{ij} matches between T_i and T_j is a real number r_{ij} such that

$$0 \leqslant r_{ij} \leqslant m_{ij} \tag{2.4}$$

and

$$r_{ij} + r_{ji} = m_{ij}. \tag{2.5}$$

If a team winning a game is awarded one and the loser zero (and each gets $1/2$ in case of a tie) then r_{ij} represents the number of points accumulated by the team T_i in the m_{ij} encounters with the team T_j. In general, each of the m_{ij} comparisons made between T_i and T_j can result in a number $0 \leqslant r \leqslant 1$ for T_i (and $1 - r$ for T_j), and then r_{ij} is the sum of these m_{ij} numbers.

A *tournament* is represented by its *result matrix*

$$R = (r_{ij})_{i,j=1}^{n}. \tag{2.6}$$

The *score* of a team T_i in the tournament (2.6) is the sum of its results in all its games:

$$s_i = \sum_{j=1}^{n} r_{ij} = r_{i1} + r_{i2} + \cdots + r_{in}. \tag{2.7}$$

The *score vector* is simply a listing of scores of all teams:

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}. \tag{2.8}$$

Ranking of a Tournament. Our goal is to find an ordering of the teams in the tournament, that is, a relation

$$T_i > T_j \tag{2.9}$$

meaning that the team T_i is better (in the tournament) than the team T_j. The relation $>$ should be *transitive*, i.e., if $T_i > T_j$ and $T_j > T_k$, then $T_i > T_k$ and we would like the ordering to be *total*, i.e., for any T_i and T_j we wish either $T_i > T_j$ or $T_j > T_i$ or $T_i = T_j$. We do, however, expect the possibility of $T_i = T_j$ for distinct i and j, i.e., the case when the teams T_i and T_j have equal ranking.

The key to successful ranking of the tournament (2.6) is to ask not just how the teams are related by the relation $>$, but to what degree. In other words, we ask not just whether T_i is better than T_j or vice versa but *how much better*: we try to assign a numerical value to the superiority of T_i over T_j (or of T_j over T_i).

To understand this idea, consider the case when the "tournament" has only two teams T_1 and T_2 and they play a large number of games, say m. It is reasonable to expect that in the long run, the better team wins more games than it loses, and the greater its superiority, the larger is the percentage of won games. If T_1 wins r games and T_2 wins the other $m - r$, then we might view the number

$$p = \frac{r}{m},$$

as the probability that T_1 wins any single game against T_2, or the probability that T_1 is better than T_2. Similarly,

$$1 - p = \frac{m - r}{m}$$

represents the probability that T_2 is better than T_1. To express this differently, the chance that T_1 beats T_2 is

$$x = \frac{p}{1-p} = \frac{r}{m-r}$$

times greater than that T_2 beats T_1. In other words, x represents the odds of T_1 winning against T_2. (Note that T_1 has an x times better record than T_2.) Note also that if T_1 wins all m games, then $p = 1$ and $x = \infty$, so x can be any number between 0 and ∞ and $0 \leqslant p \leqslant 1$.

Thus we introduce, as yet unknown, numbers

$$x_{ij}, p_{ij} \qquad i, j = 1, \ldots, n$$

with the stipulation that $0 \leqslant p_{ij} \leqslant 1$, $0 \leqslant x_{ij} \leqslant \infty$ and such that

$$x_{ji} = \frac{1}{x_{ij}} \quad \text{and} \quad p_{ji} = 1 - p_{ij}; \tag{2.10}$$

moreover, the p's and x's are related by the equations

$$x_{ij} = \frac{p_{ij}}{1-p_{ij}}, \tag{2.11}$$

$$p_{ij} = \frac{x_{ij}}{1+x_{ij}}.$$

(The second equation is obtained by solving the first equation for p_{ij}.)

We try to assign these x's (or p's) to all pairs of teams T_i, T_j. The intended meaning is that if $x_{ij} > 1$ $\left(\text{or } p_{ij} > \frac{1}{2}\right)$ then T_i is better than T_j and if $x_{ij} = 1$ $\left(\text{or } p_{ij} = \frac{1}{2}\right)$, then T_i and T_j are equal. As the intended interpretation of each x_{ij} is the relative strength of T_i and T_j, the x's should follow the laws of probabilities: for any i, j and k,

$$x_{ik} = x_{ij} \cdot x_{jk}. \tag{2.12}$$ (*)

(Since $0 \cdot \infty$ is undefined, we shall consider the equation (2.12) valid for all i, j, k whenever the right hand side is $0 \cdot \infty$). The equation (2.12) takes the following form for the associated p's (and any i, j, k):

$$p_{ij} \cdot p_{jk} \cdot p_{ki} = p_{kj} \cdot p_{ji} \cdot p_{ik}. \tag{2.13}$$

Thus we call a matrix

$$P = (p_{ij})_{i,j=1}^{n} \tag{2.14}$$

a *ranking matrix* (or simply a ranking) if the p_{ij} satisfy the equations (2.13).

Note that the equations (2.12) guarantee transitivity of the relation $T_i > T_j$: If $x_{ij} > 1$ and $x_{jk} > 1$, then $x_{ik} > 1$, and so if T_i is better than T_j and T_j is better than T_k, then T_i is better than T_k.

Since each p_{ij} is interpreted as the probability that the team T_i beats the team T_j, the expected (most probable) result of T_i in m_{ij} games against T_j is

(*)The following argument justifies the validity of the equation (2.12): Suppose we can compare the objects T_i and T_k only indirectly by comparing T_i with T_j and T_j with T_k and we do it a large number of times, say M. In $M \cdot p_{ij}$ cases, T_i looks better than T_j, and of these $M \cdot p_{ij}$ cases, T_j looks better than T_k exactly $M \cdot p_{ij} \cdot p_{jk}$ times and worse $M \cdot p_{ij} \cdot (1 - p_{jk})$ times. Whenever we find T_i better than T_j and T_j better than T_k we conclude that T_i is better than T_k, but when T_i is found better than T_j and T_j worse than T_k we reserve our judgment about T_i and T_k.

A similar situation arises in the $M \cdot (1 - p_{ij})$ cases when T_j is deemed better than T_i. Thus we have $M \cdot p_{ij} \cdot p_{jk}$ cases when T_i is declared better than T_k and $M \cdot (1 - p_{ij}) \cdot (1 - p_{jk})$ cases when T_k is considered better. It follows that

$$x_{ik} = \frac{M \cdot p_{ij} \cdot p_{jk}}{M \cdot \left(1 - p_{ij}\right) \cdot \left(1 - p_{jk}\right)} = x_{ij} \cdot x_{jk}.$$

$$m_{ij} \cdot p_{ij}$$

and the expected score of the team T_i in a tournament with schedule $M = (m_{ij})$ is

$$\sum_{j=1}^{m} m_{ij} p_{ij} = m_{i1} p_{i1} + m_{i2} p_{i2} + \cdots + m_{in} p_{in}.$$

This gives us another condition on the p_{ij}: if these numbers are the correct probabilities obtained from the tournament (2.6), then the expected score of each team has to be equal to the actual score of the team in that tournament: For each $i = 1, \ldots, n$,

$$\sum_{j=1}^{m} m_{ij} \cdot p_{ij} = s_i. \tag{2.15}$$

This reasoning leads us to the definition of the central concept of this study:

A Ranking of the Tournament

$$R = (r_{ij}) \tag{2.6}$$

is a matrix

$$P = (p_{ij}) \tag{2.14}$$

that satisfies, for all $i, j, k = 1, \ldots, n$

$$p_{ij} \cdot p_{jk} \cdot p_{ki} = p_{kj} \cdot p_{ji} \cdot p_{ik} \tag{2.13}$$

and

$$\sum_{j=1}^{m} m_{ij} \cdot p_{ij} = s_i \tag{2.15}$$

where $M = (m_{ij})$ is the schedule and $\mathbf{s} = (s_i)$ is the score (and satisfy (2.5) and (2.7)).

It is my intention to prove that the equations (2.13) and (2.15) have a solution P; i.e., a ranking exists. And, if the tournament satisfies a certain natural condition, then P is unique. The ranking matrix P then yields a linear ordering of the teams in the tournament, defined by

$$T_i > T_j \quad \text{if} \quad p_{ij} > \tfrac{1}{2} \quad (\text{or } x_{ij} > 1)$$
$$T_i = T_j \quad \text{if} \quad p_{ij} = \tfrac{1}{2} \quad (\text{or } x_{ij} = 1).$$

If $T_i > T_j$ ($T_i = T_j$), we say that the team T_i has *higher* (*equal*) *ranking* than the team T_j.

Note that the ranking P is determined by M (the schedule) and s (the score), and does not depend directly on R (the results). This is not surprising: as in a round-robin tournament, it is the total score that matters, not how the points are accumulated.

Comparability. It is clear that the tournament has to satisfy some conditions if we want to get a meaningful ranking. For instance if all $m_{ij} = 0$ (an "empty" tournament) then we have no basis for comparison of the teams. Similarly, if the tournament consists of two disjoint groups of teams with no games played between teams in different groups (like the American and National Leagues in a regular season), then we have no basis for comparison of teams in different groups. Another example is when there are two teams with a perfect score (beating all their opponents). Again, the results above do not provide a basis for comparison of the two teams. (As this situation is of considerable practical significance, one might want to use additional criteria for ranking such teams; see the remark on tie breaking in section 4.)

Definition. Teams T_i and T_j are *comparable* (in the tournament $R = (r_{ij})$) if there exists a sequence

$$i_1, i_2, \ldots, i_k \tag{2.16}$$

such that all the numbers

$$r_{i_1 i_2}, r_{i_2 i_3}, \ldots, r_{i_{k-1} i_k} \tag{2.17}$$

are positive, and either

$$i_1 = i \quad \text{and} \quad i_k = j \tag{2.18}$$

or

$$i_1 = j \quad \text{and} \quad i_k = i. \tag{2.19}$$

(Example: T_2 beats T_3, T_3 ties with T_7, T_7 beats T_4. T_2 and T_4 are comparable because $r_{2,3} = 1$, $r_{3,7} = \frac{1}{2}$, $r_{7,4} = 1$.)

If (2.18) is the case, then we use the notation

$$T_i \succcurlyeq T_j;$$

if both $T_i \succcurlyeq T_j$ and $T_j \succcurlyeq T_i$, then we say that T_i and T_j are *similar*, and denote this relation

$$T_i \equiv T_j.$$

Note that $\succcurlyeq$ is transitive and $\equiv$ is transitive and symmetric. We also let, by definition, $T_i \equiv T_i$, so that $\equiv$ is an equivalence relation. The equivalence classes of $\equiv$ are partially ordered by $\succcurlyeq$.

As an example, consider a tournament with a result matrix

	A	B	C	D
A		1	1	
B	0			1
C	0			1
D		0	0	

In this example, all teams are comparable except the pair B and C: $A \succ B \succ D$ and $A \succ C \succ D$.

The theorems.

THEOREM 1. *Every tournament in which all teams are comparable has a unique ranking.*

Precisely: If

$$R = (r_{ij}) \qquad i, j = 1, \ldots, n \tag{2.6}$$

is a tournament with schedule $M = (m_{ij})$ and score $\mathbf{s} = (s_i)$ and if each pair T_i, T_j is comparable, then there exists a unique ranking

$$P = (p_{ij}) \tag{2.14}$$

which satisfies the equations

$$p_{ij} \cdot p_{jk} \cdot p_{ki} = p_{kj} \cdot p_{ji} \cdot p_{ik} \tag{2.13}$$

and

$$\sum_{j=1}^{n} m_{ij} \cdot p_{ij} = s_i. \tag{2.15}$$

There is, in fact, a more general theorem which I will state presently. If the tournament contains incomparable teams, we can still attempt to rank the teams that are comparable.

So let us call a *partial ranking* of a tournament $R = (r_{ij})$ a partial matrix

$$P = (p_{ij})$$

in which the numbers p_{ij} are defined for those and only those pairs T_i, T_j which are comparable, and likewise satisfies the equations (2.13) and (2.15); that is, the equation (2.13) for those p's that are defined. (Note that we don't have to make qualifications about (2.15) as p_{ij} is defined when $m_{ij} \neq 0$.)

THEOREM 2. *Every tournament has a unique partial ranking.*

The third theorem is just a verification of the expected, namely that our method generalizes the simple method of comparing the win-loss record of teams with the same schedule such as in a round-robin tournament. Let us say that the teams T_i and T_j have *identical schedules* if for every $k \neq i,j$

$$m_{ik} = m_{jk}. \tag{2.20}$$

For instance, in the American League (2.3), any two eastern teams have identical schedules (and so do any two western teams).

THEOREM 3. *Let T_i and T_j be two teams with identical schedules in the tournament.*

(1) *If T_i has a higher score* (i.e, $s_i > s_j$), *then T_i has higher ranking than T_j.*
(2) *T_i and T_j have equal ranking if and only if they are comparable* and have the same score* (i.e., $s_i = s_j$).

In particular, the ranking of a round-robin tournament produces the same ordering of teams as the ordering of teams by their total scores.

We conclude this section with several remarks. One is that the uniqueness of partial ranking is best possible, in the following sense: There is no unique partial ranking that would be defined also for incomparable teams. More precisely: It can be shown (see Lemma 9 in Section 3) that for any tournament, if T_i and T_j are incomparable, then for any number p such that $0 \leqslant p \leqslant 1$ there exists a ranking $P = (p_{ij})$ of the tournament (satisfying (2.13) and (2.15)) such that $p_{ij} = p$.

Our next remark concerns the concept of comparability. Although the definition ((2.16), (2.17)) uses the result matrix (r_{ij}), the property itself depends only on the schedule matrix M and the score vector s. This claim will be proved in Section 3 (Lemma 5).

Finally, let me address the practical problem of how to compute the ranking of a given tournament. By Theorem 1, if all teams are comparable, then the tournament has a unique ranking, i.e., a unique solution $P = (p_{ij})$ of the equations (2.13) and (2.15). As shown below, this amounts to solving a system (3.2) of nonlinear equations. Our proof in Section 3 shows that the system (3.2) has a solution, by invoking the fixed point theorem, but does not indicate how to compute a solution. In practice, an approximate (but sufficiently accurate) solution can be obtained by clever successive iterations. For the examples in Section 4, a pocket calculator was a sufficient tool. For the baseball example in Appendix 2, I used a microcomputer, but for the football example in Appendix 1, I had to turn for help to an IBM 360/270 computer.

3. Proofs. The proofs in this Section are more technical than the rest of the paper. Those who are willing to believe the Theorem without reading the proof may skip Section 3 and resume reading in Section 4.

*Note that having identical schedules and the same score does not make two teams comparable. Consider the example when neither team has played any game.

Existence of a Ranking. Since Theorem 1 is a special case of Theorem 2, we shall give a proof of Theorem 2. Let n be fixed and let

$$R = (r_{ij})_{i,j}, \qquad M = (m_{ij})_{ij}, \qquad \mathrm{s} = (s_i)_i$$

be the result matrix, the schedule matrix and the score vector of a given tournament. We shall first prove that a partial ranking exists.

First we reduce the problem to the case when all teams in the tournament are in the same equivalence class given by the similarity relation $\equiv$. If E is such an equivalence class, then the restriction

$$R_E = (r_{ij})_{i,j \in E}$$

of the result matrix R describes a *subtournament* with result matrix R_E. And we have

LEMMA 1. *If for each equivalence class E of $\equiv$, the subtournament R_E has a ranking, then R has a partial ranking.*

Proof. For each equivalence class E, let $P_E = (p_{ij})_{i,j \in E}$ be a ranking of R_E. We define a partial matrix $P = (p_{ij})$ as follows: For each pair i, j such that $T_i \succcurlyeq T_j$, if $T_i \equiv T_j$, then p_{ij} is already defined; if $T_i \not\equiv T_j$, then we let $p_{ij} = 1$ (and $p_{ji} = 0$). Since $\succcurlyeq$ is transitive, it is easy to see that the equations (2.13) remain valid for the newly defined p_{ij}. As for the equations (2.15) note that when $T_i \succ T_j$ (i.e., $T_i \succcurlyeq T_j$ and $T_i \not\equiv T_j$) then $r_{ij} = m_{ij}$ (by definition of $\succcurlyeq$) and so $m_{ij} \cdot p_{ij} = r_{ij}$. It follows that the partial ranking P satisfies the equations (2.15) as well. □

So from now on we assume that any two teams in the tournament are similar:

$$T_i \equiv T_j \text{ holds for any } i, j. \tag{3.1}$$

And we want to find numbers p_{ij} satisfying (2.13) and (2.15).

LEMMA 2*. *Let $R = (r_{ij})$ be a tournament that satisfies the condition* (3.1) *and let $M = (m_{ij})$ and $\mathrm{s} = (s_i)$ be the corresponding schedule matrix and the score vector. There exist numbers $v_1, \ldots, v_n$ which satisfy the following system of equations, $i = 1, \ldots, n$:*

$$\sum_{j=1}^{n} \frac{m_{ij}}{1 + e^{v_j - v_i}} = s_i. \tag{3.2}$$

Before proving Lemma 2, let us show that a solution of (3.2) gives a ranking of R. So assume that $v_1, \ldots, v_n$ satisfy (3.2) and let, for each i and j,

$$p_{ij} = \frac{1}{1 + e^{v_j - v_i}}.$$

It is clear that the p_{ij} satisfy the equations (2.15), and noting that

$$p_{ij} = \frac{e^{v_i}}{e^{v_i} + e^{v_j}},$$

one easily verifies the equations (2.13) as well.

So our goal now is to prove Lemma 2.

Proof of Lemma 2. Let $\mathfrak{X}$ be the set of all n-tuples $\mathrm{v} = (v_1, \ldots, v_n)$ such that

$$v_1 + \cdots + v_n = 0.$$

$\mathfrak{X}$ is a closed subspace of the n-dimensional Euclidean space $\mathbb{R}^n$. We define an operator T on the space $\mathfrak{X}$. For each $\mathrm{v} \in \mathbb{R}^n$, let $F(\mathrm{v})$ be the left-hand side of (3.2); more precisely, let $F(\mathrm{v}) =$

*As the numbers x_{ij} have to satisfy (2.12), it is natural to consider their logarithms instead, thus looking for numbers v_i satisfying $v_i - v_j = \log x_{ij}$, and equivalently, $p_{ij} = 1/(1 + e^{v_j - v_i})$.

$(F_1(\mathbf{v}), \ldots, F_n(\mathbf{v}))$ where

$$F_i(\mathbf{v}) = \sum_{j=1}^{n} \frac{m_{ij}}{1 + e^{v_j - v_i}} \qquad (i = 1, \ldots, n).$$

And for each $\mathbf{v} \in \mathcal{X}$, we let

$$T\mathbf{v} = \mathbf{v} + \mathbf{s} - F(\mathbf{v}). \tag{3.3}$$

Let m denote the number of all games played in the tournament:

$$m = \frac{1}{2} \sum_{i,j} m_{ij}.$$

Note that

$$\sum_{i=1}^{n} s_i = m$$

and that for each $\mathbf{v} \in R^n$,

$$\sum_{i=1}^{n} F_i(\mathbf{v}) = \sum_{i,j} \frac{m_{ij}}{1 + e^{v_j - v_i}} = \frac{1}{2} \sum_{i,j} m_{ij} \left(\frac{e^{v_i}}{e^{v_i} + e^{v_j}} + \frac{e^{v_j}}{e^{v_i} + e^{v_j}} \right)$$
$$= \frac{1}{2} \sum_{i,j} m_{ij} = m.$$

It follows that if $\mathbf{v} \in \mathcal{X}$ and $\mathbf{z} = T\mathbf{v}$, then $\sum_{i=1}^{n} z_i = 0$ and so $T\mathbf{v} \in \mathcal{X}$. So T maps the space $\mathcal{X}$ into itself and clearly is continuous. To solve the equations (3.2) it suffices to find a fixed point of T, i.e., some $\mathbf{v} \in X$ such that

$$T\mathbf{v} = \mathbf{v}.$$

For then $\mathbf{s} - F(\mathbf{v}) = \mathbf{0}$ by (3.3) and $\mathbf{v}$ is a solution of (3.2).

To find a fixed point of T in $\mathcal{X}$ we shall use the following form of Brouwer's *Fixed Point Theorem*:

If C is a nonempty compact convex subset of R^n and if f is a continuous function on C such that $f(\mathbf{v}) \in C$ for all $\mathbf{v} \in C$, then f has a fixed point in C.

Thus Lemma 2 will follow from

LEMMA 3. *There is a nonempty compact convex subset C of $\mathcal{X}$ such that $T\mathbf{v} \in C$ for all $\mathbf{v} \in C$.*

Before we prove Lemma 3, we prove a technical lemma about F. Let d be a positive number, and let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathcal{X}$. A *gap* in $\mathbf{v}$ of length d is an (open) interval of length d such that each v_i lies outside the interval. The *top of the gap* is the set of all indices i such that v_i lies above the interval, and the *bottom of the gap* is defined similarly. If i is in the top and j is in the bottom then $v_i - v_j \geqslant d$. We always assume that both the top and the bottom of the gap are nonempty sets.

We recall that

$$m = \frac{1}{2} \sum_{i,j} m_{ij}$$

is the total number of games played. Let ε denote the least r_{ij} that is not zero:

$$\varepsilon = \min\{r_{ij} \colon i, j = 1, \ldots, n \quad \text{and} \quad r_{ij} > 0\}.$$

LEMMA 4. *Let $\mathbf{v} \in \mathcal{X}$ and let G be a gap in $\mathbf{v}$ of length $d \geqslant \log(m/\varepsilon)$. Then*

$$\sum_{i \in \text{top of } G} (s_i - F_i(\mathbf{v})) \leqslant 0.$$

Proof. Let T and B denote, respectively, the top and the bottom of the gap G. Let

$k = \sum_{\substack{i \in T \\ j \in B}} m_{ij}$ = the number of games between the top teams and the bottom teams

$l = \frac{1}{2} \sum_{i, j \in T} m_{ij}$ = the number of games between pairs of the top teams.

We made the assumption that any two distinct teams T_i and T_j are similar; it follows that there are at least one top i and one bottom j such that $r_{ji} \neq 0$, and so $r_{ij} = m_{ij} - r_{ji} \leqslant m_{ij} - \varepsilon$. Hence we have

$$\sum_{i \in T} s_i = \sum_{i, j \in T} r_{ij} + \sum_{\substack{i \in T \\ j \in B}} r_{ij} \leqslant l + (k - \varepsilon). \tag{3.4}$$

On the other hand, if $i \in T$ and $j \in B$, then $v_j - v_i \leqslant -d$, and we have

$$\sum_{i \in T} F_i(\mathbf{v}) = \sum_{i, j \in T} m_{ij} \cdot \frac{e^{v_i}}{e^{v_i} + e^{v_j}} + \sum_{\substack{i \in T \\ j \in B}} \frac{m_{ij}}{1 + e^{v_j - v_i}} \geqslant l + \frac{k}{1 + e^{-d}}. \tag{3.5}$$

Using the inequality $d \geqslant \log m/\varepsilon$, we get

$$e^{-d} \leqslant \frac{\varepsilon}{m} \leqslant \frac{\varepsilon}{k - \varepsilon}, \quad 1 + e^{-d} \leqslant \frac{k}{k - \varepsilon}, \quad \frac{k}{1 + e^{-d}} \geqslant k - \varepsilon.$$

Now, comparing (3.4) and (3.5), we conclude that

$$\sum_{i \in T} (s_i - F_i(\mathbf{v})) \leqslant 0.$$

We are now ready to complete the existence proof:

Proof of Lemma 3. Let d be a positive number such that

$$d \geqslant \log \frac{m}{\varepsilon} \quad \text{and} \quad d \geqslant m.$$

Consider the following $n - 1$ numbers

$$a_1 > a_2 > \cdots > a_{n-1} > 0$$

(where n is the total number of teams in the tournament):

$$\begin{aligned} a_{n-1} &= 2 \cdot (n-1)! \cdot d \\ a_{n-2} &= 2 \cdot (n-2)! \cdot d + a_{n-1} \\ &\cdots\cdots\cdots\cdots \\ a_3 &= 12d + a_4 \\ a_2 &= 4d + a_3 \\ a_1 &= 2d + a_2. \end{aligned} \tag{3.6}$$

Let C be the set of all $\mathbf{v} \in R^n$ such that

$$v_1 + \cdots + v_n = 0 \tag{3.7}$$

and

$$\begin{aligned} &\max\{v_i : i \leqslant n\} \leqslant a_1 \\ &\max\{v_{i_1} + v_{i_2} : i_1 \neq i_2\} \leqslant 2a_2 \\ &\cdots\cdots\cdots\cdots \\ &\max\{v_{i_1} + \cdots + v_{i_{n-1}} : \text{distinct } i_1, \ldots, i_{n-1}\} \leqslant (n-1) \cdot a_{n-1}. \end{aligned} \tag{3.8}$$

The set C is nonempty, closed, bounded (because of (3.7) and the first inequality in (3.8)), and, being the intersection of a hyperplane with a finite number of half-spaces, it is also convex. Hence it suffices to show that $T\mathbf{v} \in C$ whenever $\mathbf{v} \in C$. Let $\mathbf{v} \in C$, and let

$$z_1 \geqslant z_2 \geqslant \cdots \geqslant z_n$$

be the enumeration of the coordinates of $\mathbf{v}$ in decreasing order, and likewise, let

$$z_1^* \geqslant z_2^* \geqslant \cdots \geqslant z_n^*$$

be the enumeration of the coordinates of $T\mathbf{v}$ in decreasing order. We already know that $T\mathbf{v}$ satisfies (3.7) so we now verify that $T\mathbf{v}$ satisfies (3.8). In the arguments that follow we tacitly use the fact that for each i

$$(T\mathbf{v})_i \leqslant v_i + s_i$$

and so

$$z_1^* + \cdots + z_i^* \leqslant z_1 + \cdots + z_i + m.$$

Also note that by the assumption that $\mathbf{v} \in C$, we have $z_i \leqslant a_i$ for all $i = 1, \ldots, n-1$ and $z_n \leqslant 0$.

We start with z_1^*.

We know that $z_1 \leqslant a_1$. If $z_1 \geqslant a_2 + d$, then since $z_2 \leqslant a_2$, z_1 is the top of a gap in $\mathbf{v}$ of length $\geqslant d$, and by Lemma 4 we have $z_1^* \leqslant z_1 \leqslant a_1$. If $z_1 \leqslant a_2 + d$, then

$$z_1^* \leqslant z_1 + m \leqslant z_1 + d \leqslant (a_2 + d) + d = a_1.$$

Next we deal with z_2^*. If $z_2 \geqslant a_3 + d$, then since $z_3 \leqslant a_3$, $\{z_1, z_2\}$ is the top of a gap in $\mathbf{v}$ of length $\geqslant d$, and by Lemma 4 we have $z_1^* + z_2^* \leqslant z_1 + z_2 \leqslant 2a_2$. If $z_2 \leqslant a_3 + d$, then

$$z_1^* + z_2^* \leqslant z_1 + z_2 + m \leqslant a_1 + (a_3 + d) + d = (a_2 + 2d) + (a_2 - 2d) = 2a_2.$$

Then z_3^* is handled similarly: If $z_3 \geqslant a_4 + d$, then $\{z_1, z_2, z_3\}$ is the top of a gap of length $\geqslant d$ and we have $z_1^* + a_2^* + z_3^* \leqslant z_1 + z_2 + z_3 \leqslant 3a_3$. If $z_3 \leqslant a_4 + d$, then

$$\begin{aligned} z_1^* + z_2^* + z_3^* &\leqslant z_1 + z_2 + z_3 + m \leqslant a_1 + a_2 + (a_4 + d) + d \\ &= (a_3 + 6d) + (a_3 + 4d) + (a_3 - 10d) = 3a_3. \end{aligned}$$

We continue in this fashion, dealing successively with $z_1^*, \ldots, z_{n-1}^*$, and verifying that (3.8) is satisfied for $F(\mathbf{v})$ when $\mathbf{v} \in C$.

This shows that the continuous function T maps the set C into itself and the proof of Lemma 3 and consequently the existence part of the proof of Theorem 1 is complete. □

Uniqueness of Ranking. We shall prove that a given tournament has at most one partial ranking. As in the proof of existence we shall first reduce the general case to the case when any two teams are similar. This requires several lemmas, starting with a result promised in Section 2.

LEMMA 5. *Let $R = (r_{ij})$ and $R^* = (r_{ij}^*)$ be two result matrices that have the same schedule matrix $M = (m_{ij})$ and the same score vector $\mathbf{s} = (s_i)$. Then the comparability relations $\preccurlyeq$ and $\preccurlyeq^*$ defined by R and R^* respectively are the same. Thus comparability depends only on M and* $\mathbf{s}$.

Proof. By reasons of symmetry it suffices to show that $T_i \succcurlyeq^* T_j$ implies $T_i \succcurlyeq T_j$, for any i, j. This will follow if we show, for all i and j,

$$\text{(3.9)} \qquad \text{if } T_i \prec T_j, \quad \text{then} \quad r_{ij}^* = 0.$$

Because if $T_i \succcurlyeq^* T_j$, then by the definition of $\succcurlyeq^*$ and by (3.9) we find $i_1, \ldots, i_k$ such that $i_1 = i$, $i_k = j$, and

$$T_{i_1} \not\prec T_{i_2} \not\prec \cdots \not\prec T_{i_k}$$

and (since any two teams that have played each other are comparable) it follows that

$$T_{i_1} \succcurlyeq T_{i_2} \succcurlyeq \cdots \succcurlyeq T_{i_k}.$$

To prove (3.9), we assume by contradiction that T_{j_0} is a maximal T_j (in the partial ordering $\prec$) such that (3.9) fails. Let E be the set of all j such that $T_j \equiv T_{j_0}$.

Using the fact (2.7) that $s_j = \sum_i r_{ji} = \sum_i r_{ji}^*$ for each j, and that $r_{ji} = m_{ji} = 0$ when i and j are not comparable, we have

$$\sum_{j \in E} \sum_{T_i \preccurlyeq T_j \text{ or } T_i \succcurlyeq T_j} r_{ji} = \sum_{j \in E} \sum_{T_i \preccurlyeq T_j \text{ or } T_i \succcurlyeq T_j} r_{ji}^*$$

and consequently

$$\sum_{j \in E} \sum_{T_i \succ T_j} r_{ji} + \sum_{j \in E} \sum_{i \in E} r_{ji} + \sum_{j \in E} \sum_{T_i \prec T_j} r_{ji} = \sum_{j \in E} \sum_{T_i \succ T_j} r_{ji}^* + \sum_{j \in E} \sum_{i \in E} r_{ji}^* + \sum_{j \in E} \sum_{T_i \prec T_j} r_{ji}^*. \tag{3.10}$$

Now the first term on the left in (3.10) is 0, and by maximality of j_0 the first term on the right is also 0. The second term on both sides of (3.10) is equal to $\frac{1}{2}\sum_{i,j \in E} m_{ij}$, the number of games between the teams in E, and so

$$\sum_{j \in E} \sum_{T_i \prec T_j} r_{ji} = \sum_{j \in E} \sum_{T_i \prec T_j} r_{ji}^*. \tag{3.11}$$

However, each r_{ji} on the left of (3.11) is equal to m_{ji} (because $T_i \prec T_j$) and it follows that each $r_{ji}^* = m_{ji}$, and so $r_{ij}^* = 0$ for all $j \in E$ and all $T_i \prec T_j$, contrary to the assumption that $r_{ij_0}^* \neq 0$ for some $T_i \prec T_{j_0}$. □

LEMMA 6. *If $P = (p_{ij})$ is a partial ranking of a tournament, then*

$$p_{ij} = 0 \quad \text{if and only if} \quad T_i \prec T_j.$$

Proof. For each i and j such that $m_{ij} \neq 0$, let

$$r_{ij}^* = m_{ij} \cdot p_{ij}.$$

Since P satisfies (2.15), it follows that $R^* = (r_{ij}^*)$ is a result matrix with the same M and s. So by Lemma 5, the relation $\preccurlyeq^*$ defined by R^* coincides with $\preccurlyeq$.

First assume that $T_i \prec T_j$. There are $i_1, \ldots, i_k$ such that $i_1 = j$, $i_k = i$ and all $r_{i_1 i_2}, \ldots, r_{i_{k-1} i_k}$ are positive; moreover there is at least one t such that $T_{i_{t+1}} \prec T_{i_t}$. Therefore $T_{i_{t+1}} \prec^* T_{i_t}$, and in particular, $T_{i_{t+1}} \not\succcurlyeq^* T_{i_t}$ and so $r_{i_{t+1} i_t}^* = 0$. Since $m_{i_{t+1} i_t} = 0$, we have $p_{i_{t+1} i_t} = 0$. Now using the basic property of the p's ((2.12) or (2.13)), applied successively to $i_1, \ldots, i_k$ we conclude that $p_{ij} = 0$.

For the converse, assume that $T_i \not\prec T_j$ and that p_{ij} is defined. Then $T_i \succcurlyeq T_j$ and hence $T_i \succcurlyeq^* T_j$. There are $i_1, \ldots, i_k$ such that $i_1 = i$, $i_k = j$ and all $r_{i_1 i_2}^*, \ldots, r_{i_{k-1} i_k}^*$ are positive. Thus all $p_{i_1 i_2}, \ldots, p_{i_{k-1} i_k}$ are nonzero, and it follows (by (2.12) or (2.13) again) that $p_{ij} \neq 0$. □

LEMMA 7. *If $P = (p_{ij})$ is a partial ranking of a tournament $R = (r_{ij})$ and if E is an equivalence class* mod $\equiv$, *then $P_E = (p_{ij})_{ij \in E}$ is a ranking of the subtournament R_E.*

Proof. We know that for each i,

$$\sum_{j=1}^{n} m_{ij} p_{ij} = s_i = \sum_{j=1}^{n} r_{ij}$$

and we have to show that for each $i \in E$,

$$\sum_{j \in E}^{n} m_{ij} p_{ij} = \sum_{j \in E} r_{ij}. \tag{3.12}$$

However, if $j \notin E$ and $m_{ij} \neq 0$, then either $r_{ij} = 0$ and $p_{ij} = 0$ (if $T_i \prec T_j$), or $r_{ij} = m_{ij}$ and $p_{ij} = 1$ (if $T_i \succ T_j$), and (3.12) follows. □

Lemma 7 reduces the problem of uniqueness of ranking to a single equivalence class E. Thus the following lemma will complete the proof:

LEMMA 8. *Assume that all teams are similar in a given tournament. If* $P = (p_{ij})$ *and* $Q = (q_{ij})$ *are rankings of the tournament, then* $P = Q$.

Proof. For each i and j let

$$x_{ij} = \frac{p_{ij}}{p_{ji}}, \quad y_{ij} = \frac{q_{ij}}{q_{ji}}.$$

Since all T_i and T_j are similar, all the p's and q's are nonzero and so x_{ij} and y_{ij} are positive numbers; moreover, they satisfy the equations (2.12). Let

$$\delta_{ij} = \frac{y_{ij}}{x_{ij}},$$

$$\rho_{ij} = \frac{\delta_{ij}}{1 + \delta_{ij}}.$$

For all i, j, and k we have

$$\delta_{ik} = \delta_{ij} \cdot \delta_{jk} \tag{3.13}$$

and so $\boldsymbol{\rho} = (\rho_{ij})_{i,j}$ is a ranking matrix (i.e. it satisfies (2.13)).

Let us assume that $P \neq Q$. Then there exist i and j such that $\delta_{ij} > 1$. Since T_i and T_j are comparable, there are $i_1, \ldots, i_k$ such that $i_1 = i$, $i_k = j$ and the numbers $m_{i_1 i_2}, \ldots, m_{i_{k-1} i_k}$ are all nonzero. It follows from (3.13) that for some t, $\delta_{i_t i_{t+1}} > 1$. In other words, there exist i and j such that $\delta_{ij} > 1$ and $m_{ij} > 0$.

Let i_0 be such that there exists some j with $\delta_{i_0 j} > 1$ and $m_{i_0 j} > 0$, and that T_{i_0} has the highest ranking by $\boldsymbol{\rho}$ among T_i for all such i. So if $\delta_{i_0 j} < 1$, then $m_{i_0 j} = 0$. It follows that whenever $m_{i_0 j} \neq 0$ then $q_{i_0 j} \geqslant p_{i_0 j}$. And because for at least one j, $m_{i_0 j} \neq 0$ and $q_{i_0 j} > p_{i_0 j}$, we get

$$\sum_{j=1}^{n} m_{i_0 j} q_{i_0 j} > \sum_{j=1}^{n} m_{i_0 j} p_{i_0 j}, \tag{3.14}$$

which is a contradiction since both sides of (3.14) are equal to s_{i_0}. □

Proof of Theorem 3. Let T_i and T_j be two teams with identical schedules. First we assume that T_i and T_j are comparable.

If T_i and T_j have the same ranking (i.e., $p_{ij} = \frac{1}{2}$), then by (2.13), $p_{ik} = p_{jk}$ for all k (for which $m_{ik} = m_{jk} \neq 0$) and so

$$s_i = \sum_k m_{ik} p_{ik} = \sum_k m_{ik} \cdot p_{jk} = s_j.$$

If T_i has higher ranking $\left(p_{ij} > \frac{1}{2}\right)$ then by (2.13) again, $p_{ik} > p_{jk}$ for all k (and since $T_i \succcurlyeq T_j$ at least one m_{ik} is nonzero), so

$$s_i = \sum_k m_{ik} p_{ik} > \sum_k m_{jk} p_{jk} = s_j.$$

We complete the proof by showing that if $s_i > s_j$, then T_i and T_j are comparable. If they were not comparable, then each k with $m_{ik} \neq 0$ would satisfy simultaneously $T_k \succ T_i$ and $T_k \succ T_j$, or simultaneously $T_k \prec T_i$ and $T_k \prec T_j$, and so either $p_{ik} = p_{jk} = 0$ or $p_{ik} = p_{jk} = 1$, which would contradict $s_i \neq s_j$.

Now both parts of Theorem 3 follow. □

A final remark. We shall now prove that the uniqueness theorem is best possible:

LEMMA 9. *Let* $R = (r_{ij})$ *be a tournament and* $P = (p_{ij})$ *its unique partial ranking. Let* T_{i_0} *and* T_{j_0} *be two incomparable teams and let* $0 \leqslant p \leqslant 1$. *Then there exists a ranking* $P' = (p'_{ij})$ *extending* P *such that* $p'_{i_0 j_0} = p$.

Proof. Let E_1 and E_2 be, respectively, the equivalence classes of T_{i_0} and T_{j_0} in the relation $\equiv$. First we extend the partial ordering $\prec$ of all the equivalence classes to a linear ordering $\prec^*$ with the stipulation that if $p = 0$, then we let $E_1 \prec^* E_2$; if $p = 1$, then we let $E_1 \succ^* E_2$; and if $0 < p < 1$, then we put E_1 and E_2 together and make it one class.

As long as p_{ij} is defined we let $p'_{ij} = p_{ij}$. When $T_i \prec^* T_j$ we let $p'_{ij} = 0$ and $p'_{ji} = 1$. We also let $p'_{i_0j_0} = p$ and $p'_{j_0i_0} = 1 - p$. Finally, if $T_i \equiv T_{i_0}$ and $T_j \equiv T_{j_0}$, we define p'_{ij}, using $p'_{ii_0} = p_{ii_0}$, $p'_{i_0j_0} = p$ and $p'_{jj_0} = p_{jj_0}$, so that the condition (2.13) is satisfied.

This way we obtain a ranking P' of R extending P. □

4. Some Practical Remarks and Examples. In this section I shall give several simple examples indicating how the method works, what it takes into account, and how it compares with methods used in practice.

Example 1. A round-robin tournament of three teams.

	A	B	C	s
A		1	$\frac{1}{2}$	$1\frac{1}{2}$
B	0		1	1
C	$\frac{1}{2}$	0		$\frac{1}{2}$

Each pair has played one game; A beat B and tied with C, and B beat C. The standard method (comparing the total scores) gives the order $A > B > C$. In view of Theorem 3 this is the correct order as it coincides with the one obtained by ranking.

To find $P = (p_{ij})$ in this case we have to find a solution of the equations

$$p_{12} + p_{13} = 1.5, \quad p_{21} + p_{23} = 1, \quad p_{31} + p_{32} = 0.5, \tag{4.1}$$

satisfying (2.13).

The second equation implies that $p_{12} = 1 - p_{21} = p_{23}$ and so $x_{12} = x_{13}$ (where $x_{ij} = p_{ij}/p_{ji}$, see (2.11)). So if we let $x = x_{12}$, we have $x_{23} = x$ and $x_{13} = x_{12} \cdot x_{23} = x^2$, and so $p_{31} = 1/(1 + x^2)$ and $p_{32} = 1/(1 + x)$, and the third equation in (4.1) becomes

$$\frac{1}{1+x} + \frac{1}{1+x^2} = \frac{1}{2}$$

or

$$x^3 - x^2 - x - 3 = 0. \tag{4.2}$$

The equation (4.2) has the (approximate) solution $x = 2.13$. This number indicates that the team A is 2.13 times better than the team B and B is 2.13 times better than C. The ranking matrix is

(4.3)

	A	B	C
A		.68	.82
B	.32		.68
C	.18	.32	

Example 2. An incomplete tournament of four teams

	A	B	C	D	s
A		1		$\frac{1}{2}$	$1\frac{1}{2}$
B	0		1	1	2
C		0		1	1
D	$\frac{1}{2}$	0	0		$\frac{1}{2}$

The team A has only played two games and although its total score is less than B's score, we would expect it to get a higher ranking than B (mainly because we would expect it at least to get a tie in its missing game with C). Our method confirms this expectation.

This case would be difficult to solve algebraically; I have done it with a pocket calculator. The ranking matrix is

(4.4)

	A	B	C	D
A		.595		.905
B	.405		.725	.870
C		.275		.725
D	.095	.130	.275	

Note that (4.4) does not show the value of p_{13} as it does not enter into the calculations.

In practice, when dealing with a large number of teams, exhibiting the ranking matrix is not the most convenient *presentation* of the ranking. There are several other possible ways that I shall now discuss (and mention another one in the next example).

In the case when the tournament is an almost completed round-robin tournament as in this example, we may use the p_{ij}'s as "expected results" in the missing game, and obtain the *expected scores* of all teams. In this case, $p_{13} = .793$ and we get the following (approximate) expected scores:

A	2.3
B	2
C	1.2
D	0.5

Another way is to assign the teams certain numbers on some linear scale. For instance, since the numbers x_{ij} indicate how much better the ith team is than the jth team, we may assign the last team value 1 and then for each i we assign the ith team the unique value a_i so that $x_{ij} = a_i/a_j$ for all i and j. In our example we get

(4.5)

A	9.78
B	6.69
C	2.56
D	1

(For instance, C is 2.56 times better than D, and A is $9.78/2.56 = 3.82$ times better than C.)

More practical is to use a logarithmic scale, that is, replace the numbers in (4.5) by their logarithms. The following is the ranking of the teams in our example on the logarithmic scale base 2. The significance of base 2 is that a difference of one point on the ranking scale signifies twice as good performance.

A	3.29
B	2.74
C	1.36
D	0

(C is $2^{1.36} = 2.56$ times better than D).

Example 3. Another tournament of four teams

	A	B	C	D	W.	L.	Pct.
A		2-1	2-1		4	2	.667
B	1-2			3-0	4	2	.667
C	1-2			2-1	3	3	.500
D		0-3	1-2		1	5	.167

Here A has played three times against B, winning 2 and losing 1, etc. Both A and B have the same number of wins and losses but A should be ranked higher, because it had a more difficult playing schedule (three games against C compared with B's three games against D). This is confirmed by the actual ranking:

	A	B	C	D
A		.565	.770	
B	.430			.897
C	.230			.770
D		.103	.230	

The ranking on a logarithmic scale (base 2) is as follows:

A	3.49
B	3.12
C	1.75
D	0

indicating that A has done $2^{3.49-3.12} = 1.3$ times better than B.

Often, performance in a tournament is measured by the winning percentage. In this case, both A and B have the same percentage .667, but since their schedules are different, this is not an accurate measure of their performance. This suggests yet another way in which the results of the ranking method can be expressed: we call the *weighted percentage* of a team T_i the expected

percentage of T_i in a round-robin tournament where all the results of T_i are the expected results p_{ij}, i.e.,

$$p_i = \frac{1}{n-1} \cdot \sum_{\substack{j=1 \\ j \neq i}}^{n} p_{ij}.$$

In the present example we get:

$$p_A = \tfrac{1}{3}(.565 + .770 + .918) = .751$$

$$p_B = \tfrac{1}{3}(.435 + .721 + .897) = .684$$

$$p_C = \tfrac{1}{3}(.230 + .279 + .770) = .426$$

$$p_D = \tfrac{1}{3}(.082 + .103 + .230) = .138$$

Example 4. Example of equal ranking

	A	B	C	D
A			1	$\frac{1}{2}$
B			$\frac{1}{2}$	1
C	0	$\frac{1}{2}$		
D	$\frac{1}{2}$	0		

All four teams are comparable (similar). Teams A and B have the same schedule and the same score; similarly C and D. It follows that $p_{AB} = .5 = p_{CD}$, and for all i, $p_{iA} = p_{iB}, p_{iC} = p_{iD}$. Since

$$p_{AC} + p_{AD} = 1.5, \text{etc.}$$

the ranking is

	A	B	C	D
A		.500	.750	.750
B			.750	.750
C				.500
D				

and so $A = B > C = D$.

Comparison with Other Methods. Theorems 1 and 2 should themselves be enough to convince the reader that the ranking method presented here is the only mathematically correct way of ranking tournaments. So I will only briefly touch upon the subject of comparison of our method with the methods used in practice.

In college football, the ranking is done by opinion polls. In the Appendix, I give as an example

the ranking of the 1978 season and compare it with the two (incompatible) rankings obtained by the two opinion polls.

In major league baseball, the division champions are the teams with the greatest percentage. This is fine as long as all the teams in the same division have the same schedule (see (2.3)). If this method is used for an incomplete season as it happened in the summer of this year (1981), the ranking so obtained need not agree with the ranking given by our method (see the example in Appendix).

I shall now give an example showing that ranking by percentage is not at all a reasonable method. Consider a tournament of three teams, A, B and C. A plays B six times, winning 4 and losing 2, and plays C twice, beating it both times. Common sense dictates that A is the best team of the three, no matter how B does against C. (This is confirmed by our ranking method as shown below.) However, consider the three following cases:

	A	B	C	
A		4-2	2-0	6-2
B	2-4		11-0	13-4
C	0-2	0-11		0-13

	A	B	C	W-L
A		4-2	2-0	6-2
B	2-4		12-1	14-5
C	0-2	1-12		1-14

	A	B	C	
A		4-2	2-0	6-2
B	2-4		13-1	15-5
C	0-2	1-13		1-15

Percentage:

A	.750
B	.737
C	.067

B	.765
A	.750
C	.000

A, B	.750
C	.063

In the first case A has a better percentage, in the second case B has a better percentage, and in the third case they both have .750. The correct ranking in each case is:

	A	B	C	p
A		.679	.965	.822
B	.321		.928	.625
C	.035	.072		.054

	A	B	C	p
A		.667	1	.833
B	.333		1	.667
C	0	0		0

	A	B	C	p
A		.667	.967	.822
B	.323		.933	.628
C	.033	.067		.050

with A best in all three cases.

Tiebreaking. In some instances our method does not determine the winner of the given tournament. This is when there are two teams with a perfect record, i.e., winning all their games. In such a case, the method does not assign the two teams T_i, T_j *any* number p_{ij}, because as proved in Section 2, *every* number p_{ij} between 0 and 1 satisfies the equations. A similar case is when there are two teams which tie their game, and both win all other games. In this case they have equal ranking (tie for first place).

Since these two cases are of considerable practical significance, one might desire to have a method that would decide which of the two teams is the champion. One method that suggests itself is to declare the winner the team with a more difficult playing schedule. (Another possibility would be to take into account the number of points scored in each game. This amounts to replacing the result matrix R by a different matrix, in effect, solving a different mathematical

problem.) This can be determined either by considering the arithmetical mean of the opponents' ranking on the logarithmic scale, or by comparing the medians of the two teams' opponents, or possibly by some other, equally plausible, procedure. However, since the teams are actually incomparable (in the first case) or actually have equal ranking (in the other case), mathematics alone can't decide which tiebreaking method is preferable.

Appendix 1 The 1978 College Football Season

In the adjoining table we give the list of the top forty teams in the ranking of the 177 college football teams in Divisions I-A and I-AA in the 1978 season.

The 1978 College Football Season

Team (AP, UPI)	W	L	T	log scale	x_1	x_{20}	x_{50}	x_{100}
1. Alabama (1, 2)	11	1	0	8.833		37:1	290:1	2550:1
2. USC (2, 1)	12	1	0	8.784	1:1.03	35	280	2450
3. Penn State (4, 4)	11	1	0	7.510	2.5	15	116	1000
4. Oklahoma (3, 3)	11	1	0	7.319	2.9	13	101	890
5. Michigan (5, 5)	10	2	0	6.259	6.0	6.1	49	430
6. Notre Dame (7, 6)	9	3	0	5.810	8.1	4.5	36	310
7. Nebraska (8, 8)	9	3	0	5.804	8.2	4.5	35	310
8. Clemson (6, 6)	11	1	0	5.505	10	2.6	29	250
9. Texas (9, 9)	9	3	0	5.010	14	2.6	20	180
10. Michigan State (12, -)	8	3	0	4.828	16	2.3	18	160
11. Purdue (13, 13)	9	2	1	4.685	1:18	2.1:1	16:1	144:1
12. Arkansas (11, 10)	9	2	1	4.670	18	2	16	142
13. Houston (10, 11)	9	3	0	4.628	18	2	16	138
14. Missouri (15, 14)	8	4	0	4.541	20	1.9	15	130
15. Maryland (20, -)	9	3	0	4.151	26	1.4	11	99
16. Stanford (17, 16)	8	4	0	4.108	26	1.4	11	96
17. Georgia (16, 15)	9	2	1	3.922	30	1.2	9.6	85
18. Texas Tech	7	4	0	3.882	31	1.18	9.3	82
19. UCLA (14, 12)	8	3	1	3.785	33	1.10	8.7	77
20. Washington	7	4	0	3.641	37		7.9	70
21. North Carolina St. (18, 19)	9	3	0	3.641	1:37	1:1	7.9:1	70:1
22. Arizona State (-, 19)	9	3	0	3.530	39	1.08	7.3	65
23. Texas A & M (19, 18)	8	4	0	3.321	46	1.2	6.3	56
24. Ohio State	7	4	1	3.134	52	1.4	5.6	49
25. Iowa State	8	4	0	2.901	61	1.7	4.7	42
26. Louisiana State	8	4	0	2.646	73	2	4	35
27. Florida State	8	3	0	2.595	75	2.1	3.8	34
28. Pittsburgh	8	4	0	2.431	85	2.3	3.4	30
29. California	6	5	0	2.426	85	2.3	3.4	30
30. Southern Methodist	4	6	1	2.301	93	2.5	3.1	27
31. North Texas St.	9	2	0	2.207	1:99	1:2.7	2.9:1	26:1
32. Georgia Tech	7	5	0	1.977	116	3.2	2.5	22
33. Navy (-, 17)	9	3	0	1.775	133	3.6	2.2	19
34. Mississippi State	6	5	0	1.726	138	3.8	2.1	18
35. Auburn	6	4	1	1.698	140	3.8	2.1	18
36. Florida A & M	12	1	0	1.678	142	3.9	2	18
37. Arizona	5	6	0	1.445	167	4.6	1.7	15
38. Colorado	6	5	0	1.400	173	4.7	1.7	15
39. Tennessee	5	5	1	1.332	181	5	1.6	14
40. Tulsa	9	2	0	1.214	197	5.4	1.5	13

MAJOR LEAGUE BASEBALL, 1981 SEASON, FIRST HALF

AMERICAN LEAGUE

	SCHEDULE													
NEW YORK*		6	6	0	0	7	5	5	7	6	4	5	5	0
BALTIMORE	6		7	0	4	0	7	6	5	3	5	4	1	6
DETROIT	6	7		7	0	0	7	3	3	5	6	2	6	5
MILWAUKEE	0	0	7		7	4	6	2	6	2	7	6	4	5
BOSTON	0	4	0	7		6	4	5	5	6	1	6	5	7
CLEVELAND	7	0	0	4	6		6	7	0	4	5	3	5	3
TORONTO	5	7	7	6	4	6		5	7	3	6	2	0	0
CHICAGO	5	6	3	2	5	7	5		6	6	6	2	0	0
OAKLAND*	7	5	3	6	5	0	7	6		0	7	0	7	7
TEXAS	6	3	5	2	6	4	3	6	0		0	7	6	7
CÁLIFORNIA	4	5	6	7	1	5	6	6	7	0		0	7	6
KANSAS CITY	5	4	2	6	6	3	2	2	0	7	0		7	6
SEATTLE	5	1	6	4	5	5	0	0	7	6	7	7		4
MINNESOTA	0	6	5	5	7	3	0	0	7	7	6	6	4	

	W-L		RANKING														EXPECTED SCORE
NEW YORK*	34	22		.551	.586	.595	.604	.607	.777	.507	.516	.541	.608	.718	.726	.778	101.1
BALTIMORE	31	23	.449		.536	.545	.555	.557	.740	.456	.465	.490	.559	.675	.684	.741	92.8
DETROIT	31	26	.414	.464		.509	.519	.522	.712	.421	.430	.454	.523	.643	.652	.712	86.8
MILWAUKEE	31	25	.405	.455	.491		.509	.512	.704	.411	.421	.445	.514	.634	.643	.704	85.2
BOSTON	30	26	.396	.445	.481	.491		.503	.696	.402	.411	.435	.504	.625	.635	.696	83.7
CLEVELAND	26	24	.393	.443	.478	.488	.497		.694	.400	.409	.433	.501	.623	.632	.694	83.2
TORONTO	16	42	.223	.260	.288	.296	.304	.306		.227	.234	.252	.308	.421	.432	.501	50.3
CHICAGO	31	22	.493	.544	.579	.589	.598	.600	.773		.510	.534	.602	.712	.721	.773	100.2
OAKLAND*	37	23	.484	.535	.570	.579	.589	.591	.766	.490		.525	.593	.705	.713	.767	98.7
TEXAS	33	22	.459	.510	.546	.555	.565	.567	.748	.466	.475		.569	.684	.693	.748	94.6
CÁLIFORNIA	31	29	.392	.441	.477	.486	.496	.499	.692	.398	.407	.431		.621	.631	.693	83.2
KANSAS CITY	20	30	.282	.325	.357	.366	.375	.377	.579	.288	.295	.316	.379		.510	.579	62.7
SEATTLE	21	36	.274	.316	.348	.357	.365	.368	.568	.279	.287	.307	.369	.490		.569	61.1
MINNESOTA	17	39	.222	.259	.288	.296	.304	.306	.499	.227	.233	.252	.307	.421	.431		50.4

NATIONAL LEAGUE

	SCHEDULE											
PHILADELPHIA*		5	6	6	6	6	6	1	3	3	7	6
ST. LOUIS	5		7	5	5	4	3	5	6	7	0	3
MONTREAL	6	7		5	6	5	7	3	0	3	7	6
PITTSBURGH	6	5	5		3	7	0	6	6	5	3	2
NEW YORK	6	5	6	3		6	6	3	3	0	6	7
CHICAGO	6	4	5	7	6		3	6	7	5	3	0
LOS ANGELES*	6	3	7	0	6	3		7	6	6	6	7
CINCINNATI	1	5	3	6	3	6	7		8	5	6	6
HOUSTON	3	6	0	6	3	7	6	8		6	6	6
ATLANTA	3	7	3	5	0	5	6	5	6		8	6
SAN FRANCISCO	7	0	7	3	6	3	6	6	6	8		7
SAN DIEGO	6	3	6	2	7	0	7	6	6	6	7	

	W-L		RANKING												EXPECTED SCORE
PHILADELPHIA*	34	21		.501	.557	.586	.748	.777	.471	.485	.610	.624	.639	.667	99.0
ST. LOUIS	30	20	.499		.556	.585	.747	.777	.471	.484	.609	.623	.638	.667	98.8
MONTREAL	30	25	.443	.444		.530	.702	.735	.415	.429	.555	.569	.585	.615	89.4
PITTSBURGH	25	23	.414	.415	.470		.677	.711	.387	.399	.525	.539	.556	.586	84.3
NEW YORK	17	34	.252	.253	.298	.323		.541	.231	.241	.345	.359	.374	.404	53.5
CHICAGO	15	37	.223	.223	.265	.289	.459		.204	.213	.309	.322	.337	.365	47.2
LOS ANGELES*	36	21	.529	.529	.585	.613	.769	.796		.513	.637	.650	.665	.692	102.7
CINCINNATI	35	21	.515	.516	.571	.601	.759	.787	.487		.624	.638	.653	.681	100.5
HOUSTON	28	29	.390	.391	.445	.475	.655	.691	.363	.376		.515	.531	.562	78.8
ATLANTA	25	29	.376	.377	.431	.461	.641	.678	.350	.362	.485		.517	.548	76.3
SAN FRANCISCO	27	32	.361	.362	.415	.444	.626	.663	.335	.347	.469	.483		.531	73.4
SAN DIEGO	23	33	.333	.333	.385	.414	.596	.635	.308	.319	.438	.452	.469		68.1

*DECLARED FIRST-HALF DIVISION WINNER

The numbers in parentheses are the rankings by the Associated Press and the United Press International opinion polls. The first three columns present the win-loss-tie record. The ranking itself is presented in the next column. The numbers represent the ranking on a logarithmic scale base 2 (the scale is calibrated so as the team ranked number 60 has ranking value 0). The p's and x's can be computed from this scale. For instance

$$x_{1,2} = 2^{8.833-8.784} = 2^{0.049} = 1.034$$

and

$$p_{1,2} = \frac{x_{1,2}}{1 + x_{1,2}} = \frac{1.034}{2.034} = .508$$

For each team in the table we also give four selected "odds" x_{ij}. For the nth team we list the numbers

$$x_{n,1}, x_{n,20}, x_{n,50} \quad \text{and} \quad x_{n,100}$$

comparing the team with the teams ranked number 1, 20, 50 and 100.

Two technical remarks:

1. All the 177 teams are comparable. One team had a perfect losing record, all other teams form one equivalence class.

2. Some teams played against opponents from Divisions II or III. For the purpose of this ranking, these lower division opponents are identified and treated as one single team. This approximation introduces a slight error into the ranking of the bottom teams; it has, however, no discernible effect on the high ranked teams. [This hypothetical "average Division II/III opponent" is ranked number 155 (with ranking value -3.685). The highest ranked team that played a lower division opponent is number 46, and we have $p(46, 155) = .985$, $x(46, 155) = 67$.]

Appendix 2 The Split 1981 Baseball Season

Due to a strike in the middle of the season it was decided that the 1981 season would be divided into two halves: the games played before the strike and the games played after the strike. The teams with the best winning percentage in each division were declared first-division winners. At the time of this writing, the second half is still being played.

In the table on p. 265 we give the ranking of both leagues, based on their results in the first half of the 1981 season. In each league, we list the Eastern division teams first, in decreasing rank, followed by the Western division teams. The numbers displayed next to each team represent

(1) the schedule, i.e., the number of games played against other teams in the league;
(2) the won-lost record;
(3) the ranking; the numbers p_{ij} rounded off to three decimal places;
(4) the expected score: this is calculated from the p_{ij}, and is based on the complete schedules of both leagues as displayed in (2.3).

In three cases the highest-ranked team in its division is the declared winner. Note, however, that in American League West, the declared winner (the team with the highest win-loss ratio) is ranked second by our method.

ANSWER TO PHOTO ON PAGE 245

Kurt Gödel.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P, L. The Voices of Time: A Cooperative Survey of Man's Views of Time as Expressed by the Sciences and by the Humanities, Second Edition. Ed: J.T. Fraser. U of Massachusetts Pr, 1981, lx + 710 pp, $12.50 (P). [ISBN: 0-87023-377-8] Though dealing largely with our idea and experience of time in a cultural, psychological and philosophical sense, these essays also probe with technical accuracy the problems inherent in scientific accounts of time. GHM

General, P, L. Humor in der Mathematik. Friedrich Wille. Vandenhoeck & Ruprecht, 1982, 120 pp, DM 19,80 (P). [ISBN: 3-525-40726-2] An amusing collection of mathematical jokes, anecdotes, parodies and instructional stories, having fun with and poking fun at higher (and lower) mathematics and mathematicians. Includes some traditional and some original material including musical pieces like the Hauptsatzkantata (Fundamental Theorem of Calculus Cantata). GHM

Precalculus, T(13: 1, 2). Concepts of Algebra & Trigonometry. Howard E. Campbell. Prindle, Weber & Schmidt, 1982, x + 645 pp. [ISBN: 0-87150-332-8] Standard precalculus topics plus complex numbers, elementary linear programming, sequences, series, and counting techniques. Early emphasis on graphing. Chapter summaries and review exercises. Many "word problems." JRG

Precalculus, T(13: 1). Concepts of College Algebra. Howard E. Campbell. Prindle, Weber & Schmidt, 1982, ix + 453 pp. [ISBN: 0-87150-325-5] An attractive and lucid presentation of the standard topics of college algebra including a chapter on systems of equations and inequalities, matrices and linear programming, and a chapter on sequences, series and probability. Includes optional calculator examples and exercises and chapter reviews. JNC

Education, T(15-17: 2), P. Basic Concepts of Elementary Mathematics, Fourth Edition. John M. Peterson, James E. Smith. Prindle, Weber & Schmidt, 1982, vii + 502 pp. [ISBN: 87150-333-6] For preservice and in-service elementary teachers; this edition integrates problem solving throughout the text, adds several geometric topics and a new chapter on consumer mathematics. (First Edition, TR, May 1971; Second Edition, TR, February 1975; Third Edition, TR, March 1979.) JNC

History, P, L. The Boole-De Morgan Correspondence 1842-1864. G.C. Smith. Clarendon Pr, 1982, 156 pp, $44. [ISBN: 0-19-853183-4] Chronologically arranged, slightly edited texts of 90 letters between two of the grandfathers of modern symbolic logic. Includes thumbnail biographies of Boole and DeMorgan and introductions and commentary on the subject matter of each group of letters (which ranged widely). GHM

History, L*. Books IV to VII of Diophantus' Arithmetica. Jacques Sesiano. Transl: Qusta ibn Luqa. Sources in History of Math. & Physical Sci., V. 3. Springer-Verlag, 1982, xii + 502 pp, $72. [ISBN: 0-387-90690-8] Text, English translation, and commentary on this document extant only in a recently discovered Arabic translation. Lengthy introduction discusses the history of the Arithmetica and the Arabic text. JRG

History, L*. Descartes on Polyhedra: A Study of the De Solidorum Elementis. P.J. Federico. Sources in History of Math. & Physical Sci., V. 4. Springer-Verlag, 1982, ix + 145 pp, $36. [ISBN: 0-387-90760-2] First English translation of Descartes' general treatment of polyhedra, arrived at by analogy with plane figures. Contents include a facsimile of the manuscript with transcription, translation and commentary, both on this work and comparisons with Euler. JRG

History, P, L*. The History of Combinatorial Group Theory: A Case Study in the History of Ideas. Bruce Chandler, Wilhelm Magnus. Studies in History of Math. & Physical Sci., V. 9. Springer-Verlag, 1982, viii + 234 pp, $46. [ISBN: 0-387-90749-1] A detailed survey of the development of the theory of groups defined by generators and relations, beginning with the 1882 paper by Walther von Dyck and extending into the mid-1970's. Includes a comprehensive bibliography, an invaluable resource for anyone searching for the roots of modern group theory. Concludes with chapters

discussing in general terms the sociology of mathematics research--migration, collaboration, publication patterns. LAS

Foundations, P. Logic Colloquium ´80. Ed: D. van Dalen, D. Lascar, T.J. Smiley. Stud. in Logic & Found. of Math., V. 108. Elsevier North-Holland, 1982, x + 342 pp, $55.75. [ISBN: 0-444-86465-2] Research papers originally intended for 1980 European summer meeting of Association for Symbolic Logic in Prague, Czechoslovakia. The local organizers cancelled the meeting at the last minute due to unexplained "circumstances lying beyond the control of the organizing committee." GHM

Foundations, P. Proceedings of the Herbrand Symposium: Logic Colloquium ´81. Ed: J. Stern. Stud. in Logic & Found. of Math., V. 107. Elsevier North-Holland, 1982, xi + 384 pp, $60.50. [ISBN: 0-444-86417-2] Invited papers on impact of work and ideas of J. Herbrand (1908-1931). Contributed research papers range from set theory to proof theory and computer science. GHM

Combinatorics, P. Lecture Notes in Mathematics-829: Combinatorial Mathematics VII. Ed: R.W. Robinson, G.W. Southern, W.D. Wallis. Springer-Verlag, 1980, x + 256 pp, $18 (P). [ISBN: 0-387-10254-X] Proceedings of the Seventh Australian Conference on Combinatorial Mathematics. JRG

Number Theory, T(17: 1), S*, P, L*. Lectures on the Theory of Algebraic Numbers. Erich Hecke. Grad. Texts in Math., No. 77. Springer-Verlag, 1981, xii + 239 pp, $36. [ISBN: 0-387-90595-2] A recent translation of the author´s 1923 classic. It is yet a precise and fresh introduction to the subject. Does not include exercises. CEC

Number Theory, S(18), P. Lecture Notes in Mathematics-927: The Trace Formula and Base Change for GL(3). Yuval Z. Flicker. Springer-Verlag, 1982, xii + 204 pp, $12.50 (P). [ISBN: 0-387-11500-5] A new way of writing the non-elliptic terms of the trace formula of Arthur, Jacquet and Langlands is applied to the case of GL(3) in order to study the base change problem. CEC

Algebra, T(18: 1), S, P. The Structure of Real Semisimple Lie Groups. Ed: T.H. Koornwinder. MC Syllabus 49. Math Centrum, 1982, v + 141 pp, Dfl. 17,85 (P). [ISBN: 90-6196-239-0] A joint effort of five contributors. The principal aim is an introduction to the structure theory of real semisimple Lie groups for one who is already fairly well-versed in the general complex theory. Three chapters on groups include Cartan, Iwasawa, Bruhat decompositions, Tits systems, Furstenburg boundary. Two chapters on semisimple algebras include classification for the real case. References, index, diagrams. JS

Calculus, T?(13: 2), S. Calculus by Calculator: Solving Single-Variable Calculus Problems with the Programmable Calculator. Maurice D. Weir. Prentice-Hall, 1982, ix + 387 pp, $15.95 (P). [ISBN: 0-13-111922-2] Introductory calculus text designed to be used with the Texas Instruments TI-59 programmable calculator. Fully utilizes the calculator´s capabilities, but in a rather cookbook style. RSK

Real Analysis, T(17: 1, 2), S*, L.** Foundations of Modern Analysis. Avner Friedman. Dover Pub, 1982, vi + 250 pp, $5.50 (P). [ISBN: 0-486-64062-0] Welcome unabridged, corrected republication of the 1970 edition (TR, January 1971). Excellent buy for use as text or as reference. Self-contained treatment of Lebesgue integration. Measure, metric spaces, functional analysis in Banach spaces and spectral theory in Hilbert spaces. Applications to topics in differential equations. JK

Complex Analysis, T(17: 1), S, P, L. Starting with the Unit Circle: Background to Higher Analysis. Loo-keng Hua. Trans: Kuniko Weltin. Springer-Verlag, 1981, xi + 179 pp, $32. [ISBN: 0-387-90589-8] An introduction to harmonic analysis in several variables. The exposition is explicit and elementary. Includes discussions of the Dirichlet problem on the unit ball in n-dimensions, the Lorentz group, the axioms of special relativity, and partial differential equations of mixed type. Very few exercises. CEC

Complex Analysis, T*(15: 1, 2), L. Complex Analysis. Joseph Bak, Donald J. Newman. Undergrad. Texts in Math. Springer-Verlag, 1982, x + 244 pp, $19.80. [ISBN: 0-387-90615-0] A carefully written and illustrated text containing all the essential topics for a first course plus enough additional topics and applications for a solid one-year treatment. Abundant, well chosen exercises. TAV

Complex Analysis, T?(18: 2), P. Lectures on Riemann Surfaces. Otto Forster. Transl: Bruce Gilligan. Grad. Texts in Math., No. 81. Springer-Verlag, 1981, viii + 254 pp, $35. [ISBN: 0-387-90617-7] A formal treatment in the definition-theorem-proof format. A few exercises have been added to this translation along with minor revisions. The translation is very smooth, although the treatment is a bit lifeless and sterile. TAV

Differential Equations, S(18), P. Nonlinear Partial Differential Equations and their Applications. Collège de France Seminar, Volume III. Ed: H. Brezis, J.L. Lions. Research Notes in Math., No. 70. Pitman Pub, 1982, 432 pp, $24.95 (P). [ISBN: 0-273-08568-9] Twenty written lectures from the 1980-81 weekly Seminar on Applied Mathematics at the Collège de France. The applications are to various fields: control theory, physics, dynamical systems, etc. PZ

Differential Equations, P*. Volterra and Functional Differential Equations. Ed: Kenneth B. Hannsgen, et al. Lect. Notes in Pure & Appl. Math., V. 81. Dekker, 1982, xii + 333 pp, $45 (P). [ISBN: 0-8247-1721-X] Full contents of the 24 invited lectures at the conference on Volterra and functional differential equations held at Virginia Polytechnic Institute and State University in

Blacksburg, June 1981. Ideal guide and unique reference for current knowledge for researchers, mathematicians, and computer scientists. JK

Differential Equations, T(17: 1, 2), S*, P*. Introduction to the Theory of Linear Partial Differential Equations. Jacques Chazarain, Alain Piriou. Stud. in Math. & Its Appl., V. 14. Elsevier North-Holland, 1982, xiv + 559 pp, $81.50. [ISBN: 0-444-86452-0] Introduction to a variety of recent methods. Applications to numerous classic problems. Detailed proofs. Flexible. Suitable for introductory course as well as for advanced students and research workers. Formerly journal-dispersed results are included. JK

Differential Equations, T(16), S. Equations Différentielles. Hervé Reinhard. Gauthier-Villars, 1982, xiv + 446 pp, (P). [ISBN: 2-04-015431-0] Attempts to bridge the gaps between the classical "great works" in differential equations and introductory texts in the subject. Emphasis on stability theory and operator theory. Appropriate for a senior-level seminar or as supplementary reading. SG

Differential Equations, S(18), P. Local Bifurcation and Symmetry. A. Vanderbauwhede. Research Notes in Math., No. 75. Pitman Pub, 1982, 350 pp, $23.95 (P). [ISBN: 0-273-08569-7] A study of the influence of symmetric properties of mappings on bifurcation properties. Applications to O(2) and SO(2) symmetry, including Hopf bifurcation. An extensive bibliography of 273 references. JG

Differential Equations, P, L. The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors. Colin Sparrow. Appl. Math. Sci., No. 41. Springer-Verlag, 1982, xii + 269 pp, $19.80 (P). [ISBN: 0-387-90775-0] An informative case study in the mathematics of turbulence: a detailed, computer-assisted analysis of the equations $x' = s(y-x)$, $y' = rx - y - xz$, $z' = xy - bz$ (introduced in 1963 by E.N. Lorenz in the study of atmospheric dynamics) as an archetype of important current research into chaotic behavior and strange attractors. LAS

Analysis, T(18: 2), P. An Introduction to Ergodic Theory. Peter Walters. Grad. Texts in Math., No. 79. Springer-Verlag, 1982, ix + 250 pp, $28. [ISBN: 0-387-90599-5] A revision and expansion of Springer Lecture Notes No. 458. The material is updated and chapters on ergodic theory of continuous transformations of compact metric spaces and other recent results are added. TAV

Analysis, T(18: 3), L. Ergodic Theory. I.P. Cornfeld, S.V. Fomin, Ya. G. Sinai. Grund. der math. Wissenschaften, B. 245. Springer-Verlag, 1982, x + 486 pp, $48. [ISBN: 0-387-90580-4] In four parts: Part 1 presents elementary (but not simple) dynamical systems and investigates their ergodicity properties; Part 2 is devoted to abstract ergodic theory emphasizing entropy; Part 3 treats spectral theory; and Part 4 discusses approximation. Contains appendices and a substantial bibliography. TAV

Analysis, P. Classification Problems in Ergodic Theory. William Parry, Selim Tuncel. London Math. Soc. Lect. Note Ser., No. 67. Cambridge U Pr, 1982, 101 pp, $14.95 (P). [ISBN: 0-521-28794-4] Presumes measure and integration and a familiarity with functional analysis through Hilbert spaces, as well as the classical ergodic theory. By classifying measure-preserving transformations in terms of how they "respect" structure, the authors investigate their applications. TAV

Analysis, S(18), P. Differential Calculus and Holomorphy: Real and Complex Analysis in Locally Convex Spaces. Jean Francois Colombeau. Math. Stud., No. 64. Elsevier North-Holland, 1982, xii + 455 pp, $65 (P). [ISBN: 0-444-86397-4] Presents aspects of real and complex analysis in locally convex spaces, usually either Banach or satisfying some nuclearity assumption. Many classical results and ideas, e.g., Taylor's theorem, mean value theorem, holomorphy, pseudoconvexity, convolution, are generalized from the finite-dimensional setting, with new and different proofs. Hitherto unavailable in book form; accessible to nonspecialists. PZ

Analysis, S*(18), L. Problems in Analysis. Bernard Gelbaum. Springer-Verlag, 1982, vii + 228 pp, $28. [ISBN: 0-387-90692-4] For students with a background in real analysis, measure, topology, and topological vector spaces; 518 assorted nontrivial problems chosen from these areas followed by careful complete solutions. The third volume in this fascinating series. TAV

Algebraic Geometry, P. Lecture Notes in Mathematics-961: Algebraic Geometry. Ed: J.M. Aroca, et al. Springer-Verlag, 1982, x + 500 pp, $25 (P). [ISBN: 0-387-11969-8] Proceedings of a 1981 conference held in Spain containing 19 papers on various aspects of algebraic geometry. SG

Geometry, T(14-17: 1), L.** Transformation Geometry: An Introduction to Symmetry. George E. Martin. Undergrad. Texts in Math. Springer-Verlag, 1982, xii + 237 pp, $28. [ISBN: 0-387-90636-3] This study of the automorphisms of the Euclidean plane and space is based on elementary high school geometry but done with a mathematical thoroughness that makes this an excellent pre-requisite to a standard abstract algebra course and outstanding preparation for prospective high school teachers. Teachers manual available. JNC

Geometry, S(15-17), L. The Fifty-Nine Icosahedra. H.S.M. Coxeter, et al. Springer-Verlag, 1982, 29 pp, $12 (P). [ISBN: 0-387-90770-X] A reprinting of the 1938 edition which was originally published as No. 6 of University of Toronto Studies (Mathematical Series); consists of an enumeration and description of the polyhedra that can be derived from the five Platonic solids by stellation. JNC

Geometry, T(15-17: 1), L. The Foundations of Geometry and the Non-Euclidean Plane. George E. Martin. Undergraduate Texts in Math. Springer-Verlag, 1975, xvi + 509 pp, $24. [ISBN: 0-387-90694-0] A reprinting of an excellent geometry text (TR, October 1976; Extended Review, June-July 1977). JNC

Algebraic Topology, P. The Symplectic Cobordism Ring II. Stanley O. Kochman. Memoirs No. 271. AMS, 1982, vii + 170 pp, $10 (P).

Topology, S(18), P. Dynamical Systems on Surfaces. C. Godbillon. Transl: H.G. Helfenstein. Universitext. Springer-Verlag, 1983, 201 pp, $19.80 (P). [ISBN: 0-387-11645-1] An analysis of local behavior of singular points and periodic orbits, as well as Poincaré-Bendixson theory and direction fields on surfaces. JG

Topology, T*(15-17: 1), S, L. Surface Topology. P.A. Firby, C.F. Gardiner. Series in Math. & Its Applic. Ellis Horwood Pub, 1982, 216 pp, $54.95. [ISBN: 0-85312-483-3] An appealing intuitive approach to many sophisticated aspects of two and three dimensional geometric topology, including the Euler characteristic, complexes, graphs, vector fields, index theorem, tessellations and the fundamental group. Based on a course at Exeter University, this book is distinguished both by a wealth of diagrams and by extensive linkages among topics. LAS

Operations Research, P, L. Foundations of Decision Support Systems. Robert H. Bonczek, Clyde W. Holsapple, Andrew B. Whinston. Oper. Res. & Indus. Eng. Academic Pr, 1981, xvii + 393 pp, $29.50. [ISBN: 0-12-113050-9] Interdisciplinary and generalized approach to a theory of decision support systems. Integrates concepts and tools from various fields, particularly data base management, linguistics and artificial intelligence. JRG

Operations Research, P. Deterministic and Stochastic Scheduling. Ed: M.A.H. Dempster, J.K. Lenstra, A.H.G. Rinnooy Kan. Math. & Physical Sci. D Reidel Pub, 1982, xii + 419 pp, $48. [ISBN: 90-277-1397-9] Proceedings of NATO Advanced Study and Research Institute, 1981. Part I reviews the state of the art with respect to scheduling models and the interface between deterministic and stochastic models. Part II presents recent results and directions for further research. JRG

Operations Research, T(17: 2), P, L*. The Single Server Queue, Revised Edition. J.W. Cohen. Appl. Math. & Mech., V. 8. Elsevier North-Holland, 1982, xiv + 694 pp, $85.25. [ISBN: 0-444-85452-5] In four parts: Part 1 covers the basic ideas of stochastic processes needed for the rest; Part 2 describes the model and the variety of methods used to analyze the queue; Part 3 covers variants, bulk queues, priorities, finite waiting space, etc.; and Part 4 covers recent developments. Extensive bibliography and update on literature. A massive, definitive treatment. (First Edition, TR, February 1970.) TAV

Operations Research, T(14-15: 2). Discrete and Dynamic Decision Analysis. J.T. Buchanan. Wiley, 1982, xii + 260 pp, $36.95. [ISBN: 0-471-10130-3] Covers both decision analysis and dynamic programming. Does not assume knowledge of probability or statistics. Demonstrates methods for structuring and solution of decision problems. Modest number of exercises. JRG

Optimization, T(15-16). Méthodes explicites de l'optimisation. Jean-Pierre Aubin, Pierre Nepomiastchy, Anne-Marie Charles. Dunod, 1982, 287 pp, 200 FF. [ISBN: 2-04-015402-7] Extensive treatment of quadratic programming with applications to economics; introduction to convex analysis. JRG

Probability, P. The Multiple Stochastic Integral. David Douglas Engel. Memoirs No. 265. AMS, 1982, v + 82 pp, $5 (P). The author's doctoral thesis (Yale 1979). The main result represented is that the multiple stochastic integral as introduced by Ito is a countably additive extension of the $L^2\Omega$-valued set function $X^{(n)}$. TAV

Probability, T(18: 2), P*. Lectures from Markov Processes to Brownian Motion. Kai Lai Chung. Grund. der math. Wissenschaften, B. 249. Springer-Verlag, 1982, viii + 239 pp, $34. [ISBN: 0-387-90618-5] An impressive piece of work. Chung begins with the familiar properties of Markov processes, extends and generalizes through martingales, Feller and Hunt processes to Brownian motion, ending with areas of future study. Challenging but rewarding reading. TAV

Probability, P. Markov Random Fields. Yu. A. Rozanov. Transl: Constance M. Elson. Springer-Verlag, 1982, ix + 201 pp, $42. [ISBN: 0-387-90708-4] In an attempt to formulate the Markov property for generalized random functions of several variables, the author introduces and develops the theory of random fields on continuous time domains. TAV

Statistics, P. Lecture Notes in Statistics-10: Fitting Linear Models: An Application of Conjugate Gradient Algorithms. Allen McIntosh. Springer-Verlag, 1982, vi + 200 pp, $12 (P). [ISBN: 0-387-90746-7] Presents theory and illustrates by examples (with computer output) the use of conjugate gradient algorithms to fit a variety of linear models. These algorithms require much less computer memory than more commonly used algorithms and so are more appropriate for small computers. RSK

Statistics, T(15-16: 2). Elements of Probability and Mathematical Statistics. Frederick H. Steen. Duxbury Pr, 1982, xiii + 505 pp. [ISBN: 0-87872-299-8] Fairly traditional text with a broad coverage of theoretical topics. Includes chapters on Bayesian inference and nonparametric statistics. RSK

Statistics, T*(16-17: 1, 2). Linear Models: An Introduction. Irwin Guttman. Wiley, 1982, x + 358 pp, $36.95. [ISBN: 0-471-09915-5] In the Wiley Series in Probability and Mathematical Statistics. Clearly written theoretical presentation of the "essential concepts of regression, least squares, and linear models." Presumes courses in statistics and linear algebra. RSK

Computer Programming, S, P, L*. Basic BASIC-English Dictionary for the Apple, PET and TRS-80. Larry Noonan. Dilithium Pr, 1982, 150 pp, $10.95 (P). [ISBN: 0-918398-54-1] A convenient aid for translating among the three Basic languages used by Apple II, PET and TRS-80. Part I is a dictionary of commands, functions, etc., giving English definitions and the corresponding terms in each of the three dialects; Part II is a summary of terms; appendices explain similarities and differences of ASCII codes, key words, graphics and Boolean operators. JNC

Computer Programming, S(13-14). Some Common BASIC Programs: IBM Personal Computer Edition. Lon Poole, Mary Borchers, Peter M. Burke. Osborne/McGraw-Hill, 1982, vii + 212 pp, $14.99 (P). [ISBN: 0-931988-83-7] A collection of 76 practical programs written in Basic for the I.B.M. Personal Computer; includes programs in consumer mathematics, elementary statistics, matrix calculations and numerical integration. JNC

Computer Science, T(17: 1), S. Algorithmic Language and Program Development. F.L. Bauer, H. Wössner. Texts & Mono. in Comp. Sci. Springer-Verlag, 1982, xvi + 497 pp, $29. [ISBN: 0-387-11148-4] This book presents a formal description of the fields of problem solving and computer programming. Rather than looking at a particular language and its accompanying syntax, it takes a language-independent, algorithmic approach to programming. It investigates the mathematical foundations of such common language constructs as subprograms, variables, data types, and control structures. It assumes a very high level of mathematical sophistication on the part of the reader, especially in the areas of logic, algebra, and the predicate calculus. MS

Control Theory, T(17-18: 1). Introduction to Optimal Control Theory. Jack Macki, Aaron Strauss. Undergrad. Texts in Math. Springer-Verlag, 1982, xiii + 165 pp, $24. [ISBN: 0-387-90624-X] Emphasis on motivation and explanation with many examples. One simple canonical example (the rocket car) is carried through the text. Discussion includes linear autonomous time-optimal control problems, existence theorems, and necessary conditions for optimal controls. Modest number of exercises. JRG

Applications (Actuarial Science), P?. Limits of Insurability of Risks. Baruch Berliner. Prentice-Hall, 1982, x + 118 pp, $14.95. [ISBN: 0-13-536789-1] Mostly non-mathematical discussion of what risks are still insurable. RSK

Applications (Biology), P. Lecture Notes in Medical Informatics-17: Biomedical Images and Computers. Ed: J. Sklansky, J.-C. Bisconte. Springer-Verlag, 1982, vii + 329 pp, $21 (P). [ISBN: 0-387-11579-X] Proceedings of United States-France Seminar on Biomedical Image Processing. Papers on microscopic image analysis, radiological image analysis, tomography, image processing technology. JRG

Applications (Engineering), T(16-17: 1), P.** Problems in Structural Analysis by Matrix Methods. P. Bhatt. Construction Pr, 1981, 465 pp, $22.50 (P). [ISBN: 0-86095-881-7] Unified approach via matrix methods. Digital-computer-oriented. Numerous examples with complete solutions paralleling steps in a computer program. Covers elastic analysis, elastic-plastic analysis, elastic stability plus natural frequency calculations. Presupposes a first course in structural analysis. Exercises with answers. JK

Applications (Engineering), T*(17). Variational Methods in Theoretical Mechanics, Second Edition. J.T. Oden, J.N. Reddy. Universitext. Springer-Verlag, 1983, xi + 309 pp, $20 (P). [ISBN: 0-387-11917-5] Revised and corrected edition (First Edition, TR, June-July 1977). Out-of-date sections deleted. Textbook for graduate students in mathematics and engineering science. Covers essentials of modern variational theory. JK

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

The Mathematical Association of America
The Sixty Sixth Annual Meeting of the Association
Denver, Colorado

The Sixty Sixth Annual Meeting was held in Denver, Colorado during the period January 4-9, 1983. There were 2326 registrants including 1227 members of the Association. The Program Committee consisted of: C.E. Burgess; Ronald L. Graham; Ruth I. Hoffman; Kenneth I. Gross; John H. Hodges; Gary Bitter and Stanley P. Gudder. The talks for which abstracts were submitted consisted of the following:

MAA/NCTM Panel Discussion: Please Professor, Let's Work Together on This! William E. Briggs, William Cox, Helen Holgren, Terry Kret, Billy E. Rhoades, Ronald Schnackenberg, Donald B. Small.

A survey by the MAA Committee on Secondary School Lectures has shown that there are approximately one hundred and twenty existing lectureship programs, suggesting that there is substantial interest among college/university mathematics faculty members in such programs. It is likely that most faculty members see these programs as a way to interact with high school mathematics teachers and students, i.e., to help and support high school teachers in their efforts to stimulate interest in mathematics among their students. On the other hand, most of the existing programs are very small. This suggests that high school teachers do not share the interest and sense of importance of lectureship programs that are found among college faculty. It also suggests that there may be other ways in which college and high school mathematics teachers can establish a two-way communication and support system that will benefit the high school students. The latest NRC study entitled "The State of School Science" delineates a number of serious problems with the present situation in high school mathematics. The purpose of this session was to hear directly from a panel of high school and college mathematics teachers about their perspectives on the problems to be faced and then discuss, in small groups, the issues raised by the panel members.

Nonassociative Algebras, the First 101 Years, by Marvin L. Tomber, Michigan State University.

In 1847 Cayley published the equation (i(a) i(b)) i(y) = i(a) (i(b) i(y)). This is the first public announcement of the existance of a nonassociative product. The genesis of the Cayley-Graves numbers is traced from Hamilton's treatment of the complex numbers. After the beginning of a generalized concept of algebra the study of associative and Lie algebras was initiated. In the early and mid 1930's, Zorn and Moufang began their study of alternative algebras and Jordan initiated his investigations which would lead to Jordan algebras. Early work on alternative and Jordan algebras is described along with other developments. The 101 years end in 1948 with Albert's paper on power-associative rings which played an important role in subsequent developments.

How to Throw Small Matrices Away, or, Just What Did Brown, Douglas, And Fillmore Do in 1973?, by Paul R. Halmos, Indiana University.

Operator theory can be viewed as a generalization of finite-dimensional linear algebra. The most famous theorem of linear algebra is the diagonalizability of real symmetric matrices--a result whose infinite-dimensional extension encounters many difficulties, both technical and conceptual. The purpose of this lecture is to place the familiar theorem in its proper algebraic context, extend the context to infinite-dimensional spaces, and then "throw small matrices away" (that is, form the appropriate quotient algebras). What is left is one of the biggest operator-theoretic steps forward in the 1970's, namely the celebrated theorem of Brown, Douglas, and Fillmore.

Nerve Conduction and Cardiac Fibers: Some Qualitative Problems in Differential Equations, by Jane Cronin Scanlon, Rutgers University.

The Hodgkin-Huxley equations of nerve conduction and the Noble equations for the cardiac Purkinje fiber were described and compared. Mathematical and physiological reasons were given for studying these equations from the singular perturbation viewpoint, especially studying discontinuous solutions of the corresponding degenerate systems. Results for the Noble equations were obtained by using this viewpoint and applying work of Mishchenko and Rozov. Unsolved mathematical questions which arise were described.

MAA Workshop: High Level Languages--Why PASCAL?, by John L. Van Iwaarden, Hope College.

The high-level computer language PASCAL is increasing dramatically in popularity as a first language for students in mathematics and computer science as well as engineering and other physical sciences. The literature now abounds with references to PASCAL. Mathematicians are well directed to learn something of the unique facets of this language. A workshop entitled "High Level Languages--Why PASCAL?" was presented by Professor John L. Van Iwaarden of Hope College to explain the syntax structure of the language and detail some of its unique data structures and its versatility. Comparisons of PASCAL with BASIC, COBOL, and FORTRAN were given.

A Systematic Method for Teaching Mathematical Proofs, by Daniel Solow, Case Western Reserve University.

An effective and systematic approach to teaching mathematical proofs and theoretical reasoning was presented. The approach is based on the observation that all proofs repeatedly make use of a

limited number of different proof techniques, each of which was described. Then it was shown how a proof can be explained as a sequence of applications of the individual techniques, and how a correct technique can often be chosen based on certain key words appearing in the problem under consideration. Benefits of using the proposed method, and the issue of implementing this material into current curricula was addressed.

Mathematical Modeling in Petroleum Reservoir Simulation, by Richard E. Ewing, University of Wyoming and Mobil Research and Development, Dallas, Texas.

Mathematical questions arising in the numerical simulation of enhanced recovery processes are discussed. A model for describing miscible displacement in porous media is given by a coupled system of nonlinear partial differential equations for the concentration of the invading fluid and the pressure of the total fluid. Mixed finite element methods are presented for the simultaneous approximation of the pressure and the darcy velocity from an elliptic equation with point sources and sinks. The concentration of the invading fluid is described by a nonlinear transport dominated parabolic equation with point sources which has many properties of first order hyperbolic equations.

Homoclinic Bifurcation and Phase Transition in Time Discrete Dynamical Systems (with computer pictures), by Heinz-Otto Peitgen, University of Bremen and University of Utah.

The author discusses the scenario of homoclinic bifurcation for time discrete dynamical systems, which are given by one-parameter families of area preserving diffeomorphisms of the plane. For specially selected models which are derived from numerical approximations of nonlinear elliptic boundary value problems he gave some evidence of its occurrence. He indicated how these structural changes govern the asymptotic fate of periodic orbits and initiate phase transitions in the global dynamics. Particular attention was given to the phenomenon of homoclinic bifurcation as such in the space of all dynamical systems and its connection to the occurrence of spurious solutions in numerical approximation schemes. The results were illustrated by computer-generated phase portraits. Subsequently a computer-generated movie was shown, which was produced at the University of Utah in Salt Lake City with the support of the Departments of Computer Science and Mathematics.

VisiCalc and Mathematical Algorithms: Mathematical Applications of an Electronic Spreadsheet, by Deane E. Arganbright, Whitworth College.

VisiCalc is a computer software package designed as an "electronic spreadsheet" for business and economic modeling and forecasting. This talk presents VisiCalc as a creative and dynamic tool for doing and teaching mathematics. Algorithms which are recursive, iterative, or tabular in nature can be easily implemented on VisiCalc in a way that allows the user to change parameters, initial values and constants, and instantly see the result of the changes. Examples are drawn from numerical analysis, linear programming, linear algebra, number theory, differential equations, combinatorics, statistics, calculus, and elementary mathematics.

The Role of Microcomputers in the Mathematics Curriculum, by Ruth Hoffman, University of Denver.

Computer Literacy is the fundamental issue of the decade. It has developed in the past 15 years in the secondary school from computer science courses through microcomputers used in mathematics, hence to microcomputer usage in mathematics, science and business education and in the '80's to microcomputers used in all areas of the pre-college curriculum. Colleges have moved more slowly. In a survey in the spring of 1982, 14,132 pre-college schools across the nation were officially listed as having microcomputers in their curriculum. The conclusion of the survey was that computers have definitely penetrated the education system marking the beginning of major curricular changes. A survey was also taken of 100 colleges from a list of 3,308. The conclusion for colleges is that computer literacy is ill-defined and that different institutions have widely diverse needs.

Progress Report of the National Science Board Commission on Precollege Preparation in Mathematics, by Katherine P. Layton, Beverly Hills High School.

In response to the current decline in the quality and quantity of precollege mathematics and science education in the United States, the National Science Board established the Commission on Precollege Education in Mathematics, Science and Technology. The purpose of the Commission is to define a national agenda for improving mathematics and science education. A progress report was given of the work done by the Commission during its first eight months. Included was a discussion of the problem statement, "Today's Problems, Tomorrow's Crises;" the Atlanta Public Hearings; Access to High Quality Education in Mathematics, Science and Technology for All Citizens; and studies instituted by the Task Groups on Education, Governments, Facilitators, and Recipients.

Also included on the program were lectures entitled "Applications in the Undergraduate Curriculum" by Solomon Garfunkel, and "Geometric Structures for 3-Manifolds with Symmetry" by William P. Thurston. The Association's Committee on Corporate Members sponsored a session entitled "Mathematics Publishing, Copyright, and Software." Finally, there was a session organized by Irwin Kra on behalf of the MAA/AMS Committee on Employment and Educational Policy entitled "Freshman Mathematics: Are There Alternatives to the Calculus?"

Special Sessions

Mini-Courses.

The Association sponsored five mini-courses entitled "Placement Testing," "Introduction to Microcomputers in Mathematics Instruction," "An Introduction to the Mathematical Foundations of Computer Graphics," and "Uses of Computers in Undergraduate Mathematics Instruction." Persons involved in these courses and from whom additional information can be requested are Richard H. Prosl, College of William and Mary; Klaus E. Eldridge or Donald O. Norris, Ohio University; Gerald J. Porter, University of Pennsylvania; David A. Smith, Duke University.

Contributed Papers.

For the first time at an Annual Meeting, the Association sponsored sessions for contributed papers. There were two such sessions entitled "Discrete Mathematics in the Undergraduate Curriculum Program" organized by Anthony Ralston, and "Computers in Undergraduate Mathematics Instruction" organized by Ronald Wenger. The list of presentors follows: "Proof by Algorithm," Stephen B. Maurer; "Some Elementary Applications of the Principle of Inclusion and Exclusion," Kenneth P. Bogart; "An Alternative to Calculus for Liberal Arts Students," Duane Bailey; "Some Finite Problems Neither My Computer Nor I Can Solve," Richard V. Andree; "Formal Reasoning in Discrete Mathematics," Susanna S. Epp; "Calculus with a Function Preprocessor," R.J. Knill; "3-D Image Generation in Vector Calculus," Michael Frantz; "A Hidden-Line Algorithm for Three-Dimensional Computer Graphics," William F. Rich; "APPLE Graphics Enhanced With an Attached Processor," Roy Myers; "The School of Soft Knocks," Richard L. Eisenman; "Using Microcomputers to Illustrate Concepts in Probability and Statistics," Elliot A. Tanis; "Some APPLE Computer Programs for Combinatorics Class," Donald R. Snow; "A Mathematical Analysis of the Game of Jai Alai," Louise E. Moser; "Applying the Computer to Abstract Algebra," George M. Whitson; "Computer Generated Stereo Slides-Movies-Models," Cliff Long and Thomas Hern; and "Computer Scaling of Tests," Clifton A. Lando.

The Local Arrangements Committee consisted of: Jerrold W. Bebernas; William S. Dorn; Gary W. Grefsrud; Raymond R. Gutzman; Darel W. Hardy; Zenos Hartvigson; Frieda K. Holley; Arne Magnus; Richard Osborne; Arlan B. Ramsay; William N. Reinhardt; Nancy Warren Townsend.

Board and Business Meeting

The Board Governors met at 9:00 A.M. on January 6, 1983 in the Forum of the Executive Tower Inn. The major items of business will be announced to the membership in FOCUS.

The business meeting of the Association was held at 9:30 A.M. on January 8, 1983. At this meeting, Alice Beckenbach accepted the Award for Distinguished Service to Mathematics on behalf of her late husband Edwin F. Beckenbach. The citation that accompanied the Award was read by Professor Ivan Niven, MAA President-Elect. This citation will be published in the Monthly.

The Certificate of Merit was presented to H. Hope Daly for her work as Head of Meetings Arrangements. The Secretary read a citation and the audience joined in an ovation for Ms. Daly.

The revisions to the By-Laws made necessary by offering members the choice of any one journal published by the Association were approved.

The Lester R. Ford Awards for expository Monthly articles were presented to:

Philip Davis, Brown University
"Are There Coincidences in Mathematics?"
American Mathematical Monthly

R. Arthur Knoebel, New Mexico State University
"Exponentials Reiterated."
American Mathematical Monthly

Election of Members

At its meeting on January 6, 1983, the Board elected to membership 802 applicants for individual membership and 8 applicants for academic membership. The latter follow: Hampden-Sydney College; Iowa Wesleyan College; University of Maine at Machias; University of Wisconsin at Green Bay; Drake University; Moorpark College; Liberty Baptist College; and Hamline University.

Respectfully submitted,

David P. Roselle
Secretary

CONJECTURES ON THE CRITICAL POINTS OF A POLYNOMIAL

MORRIS MARDEN
Department of Mathematics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

1. Introduction. Let us suppose that $p(z)$ is an nth degree polynomial which has *all* its zeros in the unit disk $|z| \leqslant 1$; then *all* the critical points (zeros of the derivative) of $p(z)$ also lie in the same disk $|z| \leqslant 1$. This is in fact the well-known theorem which was implied in a note of Gauss (1836) and stated and proved explicitly by Lucas (1874) (Marden [6], p. 22).

Now, instead of considering the relative positions of *all* the zeros and *all* the critical points of $p(z)$, let us choose any one zero z_0 of $p(z)$ and ask: At most how far from z_0 does the nearest critical point lie? A possible answer to this question is given by the following:

CONJECTURE I. *If $p(z)$ is an nth degree polynomial having all its zeros in the unit disk $|z| \leqslant 1$ and if z_0 is any one such zero, then at least one critical point of $p(z)$ lies on the disk $|z - z_0| \leqslant 1$.* (*Fig.* 1.)

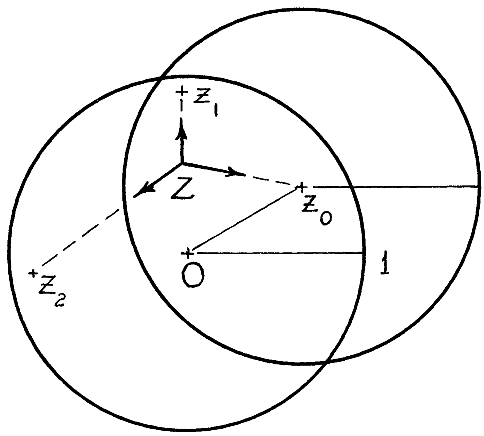


FIG. 1

This conjecture was included in the collection of *Research Problems in Function Theory*, published in 1967 by Professor Hayman [4]. Since it had been brought to Hayman's attention by Professor Ilyeff, it became known as "Ilyeff's Conjecture." Actually, Conjecture I was due to the Bulgarian mathematician B. Sendov who had acquainted me and probably others with it in 1962 at the International Congress of Mathematicians held in Stockholm.

After publication of the Conjecture, efforts were initiated by a number of very competent mathematicians to test the validity of the conjecture. Fourteen papers were published on the conjecture between 1968 and 1978. Some of these papers dealt with special cases or used various ad hoc devices. Our purpose here is to report primarily upon some general methods and results pertaining to Conjecture I.

Morris Marden received his Ph.D from Harvard under Joseph L. Walsh in 1928. As a National Research Fellow from 1928 to 1930 he did postdoctoral work at Wisconsin under E. B. Van Vleck, at Princeton under Einar Hille, at Zurich under George Pólya and at Paris under Paul Montel. Later he studied applied mathematics at Brown under Stefan Bergman. Since 1930 he has been on the University of Wisconsin-Milwaukee faculty; since 1975 as UWM Distinguished Professor Emeritus. During 1975–77 he served as Visiting Distinguished Professor at California Polytechnic State University in San Luis Obispo. He has served as a MAA visiting lecturer and has lectured abroad extensively. He has two sons: Albert, professor of mathematics at the University of Minnesota; and Philip, a pediatrician practicing in Oconomowoc, Wisconsin.

The Conjecture is obviously true for $p(z) = z^n - 1$ and polynomials of the form $p(z) = a_0 + a_2 z^2 + \cdots + a_n z^n$ where $a_0 a_n \neq 0$. For these polynomials $p'(0) = 0$ and so every zero is within unit distance from the origin, which is a critical point.

The Conjecture may be interpreted physically if we recall Gauss' Theorem that the critical points of a polynomial, which are not multiple zeros of the polynomial, are the equilibrium points in a certain force field. The field is due to particles placed at the zeros of the polynomial, the particles having masses equal to the multiplicity of the zeros and attracting with a force inversely proportional to the distance from the particle. These critical points cannot therefore be too close to any one zero since the force due to the particle at the zero would be relatively large. On the other hand, the Conjecture would imply that, if all the particles are of unit mass and situated on the disk $|z| \leqslant 1$, then at least one equilibrium point will lie within unit distance of each particle.

The Conjecture may also be given a geometric interpretation. It is known (Marden [**6**, p. 9] that the critical points ζ_1 and ζ_2 ($\neq z_1, z_2, z_3$) of the polynomial

$$p(z) = (z - z_1)^{m_1}(z - z_2)^{m_2}(z - z_3)^{m_3}$$

lie at the foci of the ellipse which touches the line segments (z_1, z_2), (z_2, z_3), and (z_3, z_1) in the points that divide these segments in the ratios m_1/m_2, m_2/m_3 and m_3/m_1 respectively. The Conjecture implies therefore that, if the vertices of this triangle z_1, z_2, z_3 all lie in the unit disk $|z| \leqslant 1$, each vertex is within unit distance from one of the foci of the inscribed ellipse (Fig. 2). A similar interpretation can be given for the critical points of the polynomial

$$p(z) = (z - z_1)(z - z_2) \cdots (z - z_n)$$

as foci of a curve Γ of class n (Marden [**6**, p. 11]), which is tangent to the sides of the polygon with vertices at the points z_j, $j = 1, 2, \ldots, n$. If true, the Conjecture would imply that within unit distance of each vertex lies at least one focus of Γ.

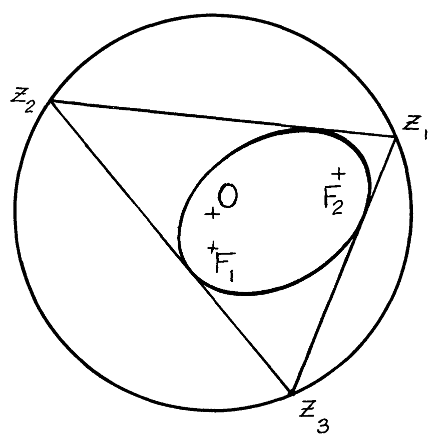


FIG. 2

2. Existence of Extremal Polynomials. Let us reformulate Conjecture I as in Phelps-Rodriguez [**16**] by study of the extremal members in the family $\mathcal{P}$ of all nth degree polynomials whose zeros all lie on the disk $|z| \leqslant 1$. Let us write $p(z)$ and its derivative $p'(z)$ in the form

$$p(z) = \prod_{j=1}^{n} (z - z_j), \quad p'(z) = n \prod_{k=1}^{n-1} (z - \zeta_k) \tag{2.1}$$

and denote by $D(a, r)$ the disk $|z - a| \leqslant r$. Let us also define

$$I(z_j) = \min\{|z_j - \zeta_k|, k = 1,2,\dots, n-1\}; I(p) = \max\{I(z_j), j = 1,2,\dots, n\};$$
$$I(\mathcal{P}) = \sup\{I(p), p \in \mathcal{P}\}.$$

According to Lucas' theorem (Marden [**6**, p. 22] any circle containing all the zeros of a polynomial $p(z)$ also contains the critical points of $p(z)$. It follows that $I(\mathcal{P}) \leqslant 2$. The example $p(z) = z^n - 1$ has $I(p) = 1$ and thus shows that $I(\mathcal{P}) \geqslant 1$. However, Conjecture I asserts that $I(\mathcal{P}) \leqslant 1$.

Any polynomial $p^* \in \mathcal{P}$ such that $I(p^*) = I(\mathcal{P})$ is called an *extremal polynomial* of $\mathcal{P}$. Let us first determine that $\mathcal{P}$ contains at least one extremal polynomial p^*.

By the definition of extremal polynomial, $\mathcal{P}$ contains a sequence $\{p_k\}$ of polynomials such that

$$\lim_{k\to\infty} I(p_k) = I(\mathcal{P}). \tag{2.2}$$

Now, $\mathcal{P}$ is a "normal family of functions" since for an arbitrary but fixed $R > 0$, all polynomials $p \in \mathcal{P}$ have, according to (2.1), the property $|p(z)| \leqslant (R+1)^n$ for $|z| \leqslant R$ and hence are uniformly bounded. This means that a subsequence of the $\{p_k\}$ can be found that in $|z| \leqslant R$ converge uniformly to nth degree polynomial $f(z) = (z - Z_1)(z - Z_2)\cdots(z - Z_n)$ and simultaneously the p_k' converge uniformly to the derivative of $f(z)$:

$$f'(z) = n(z - W_1)(z - W_2)\cdots(z - W_{n-1}).$$

By Hurwitz' Theorem (Marden [**6**, p. 4]) the zeros of $f(z)$ and $f'(z)$ are the limits of the zeros of p_k and p_k' respectively. Hence, we infer that $|Z_j| \leqslant 1$, $j = 1,2,\dots, n$; $f(z) \in \mathcal{P}$, and that $I(f) \leqslant I(\mathcal{P})$.

To show in fact that $I(f) = I(\mathcal{P})$, let us assume on the contrary that $I(f) < I(\mathcal{P})$ and write

$$I(f) = I(\mathcal{P}) - 4\varepsilon, \quad \varepsilon > 0. \tag{2.3}$$

Let us choose $\delta, 0 < \delta < \varepsilon$, so that $f(z) \neq 0$ for $0 < |z - Z_j| < \delta$, $j = 1,2,\dots, n$. If k is sufficiently large, only one zero z_{kj} of $p_k(z)$ lies in disk $D(Z_j, \delta) \subset D(Z_j, \varepsilon)$. From (2.3) we conclude that at least one zero W_w of $f'(z)$ satisfies the inequality

$$|Z_j - W_\nu| \leqslant I(f) = I(\mathcal{P}) - 4\varepsilon. \tag{2.4}$$

Choose δ so that also $f'(z) \neq 0$ for $0 < |z - W_\nu| < \delta$ and thus for k sufficiently large only one zero $\zeta_{k\nu}$ of p_k' lies in $D(W_\nu, \varepsilon)$. From

$$z_{kj} - \zeta_{k\nu} = (z_{kj} - Z_j) + (Z_j - W_\nu) + (W_\nu - \zeta_{k\nu})$$

it follows that

$$\begin{aligned} |z_{kj} - \zeta_{k\nu}| &\leqslant |z_{kj} - Z_j| + |Z_j - W_\nu| + |W_\nu - \zeta_{k\nu}| \\ &\leqslant \delta + I(\mathcal{P}) - 4\varepsilon + \delta. \end{aligned}$$

Hence, $\lim_{k\to\infty} I(p_k) \leqslant I(\mathcal{P}) - 2\varepsilon$ which contradicts (2.2). Therefore $I(f) = I(\mathcal{P})$ and hence $f(z) \equiv p^*(z)$, an extremal member of $\mathcal{P}$.

3. Properties of Extremal Polynomials. We now show that a necessary condition on each extremal polynomial $p^*(z)$ is that at least one of its zeros Z_j lies on the unit circle $C(0,1)$. This condition is to be expected in view of the fact that each zero W_k of $p^{*\prime}(z)$ is an analytic function of the zeros Z_j of $p^*(z)$ due to the equation

$$p^{*\prime}(W_k)/p^*(W_k) = \sum_{j=1}^{n} \left[1/(W_k - Z_j)\right] = 0.$$

Since a neighborhood of any Z_j with $|Z_j| < 1$ would be mapped onto a neighborhood of W_k, a small increment on Z_j in a suitable direction should increase $I(Z_j)$ contrary to the hypothesis that $p^*(z)$ is extremal. Hence, we may expect at least one Z_j to lie on circle $C(0,1)$.

That $p^*(z)$ actually has at least one zero on the unit circle is proved in Phelps-Rodriguez [**16**] as

follows. Assuming on the contrary that $|Z_j| < 1$, $j = 1,2,\ldots,n$, and labeling Z_j so that $I(Z_1) = I(p^*)$, we note that by Lucas' Theorem $|W_k| < 1$ and that $|Z_1 - W_k| \geqslant I(Z_1)$ for $k = 1,2,\ldots,n-1$. Let us choose a sequence $\zeta_{k\nu}$, $\nu = 1,2,\ldots$, with $|\zeta_{k\nu}| < 1$, $\lim_{\nu\to\infty}\zeta = W_k$, and $|Z_1 - \zeta_{k\nu}| > I(Z_1)$. Then the polynomials

$$f_\nu(z) = n(z - \zeta_{1\nu}) \cdots (z - \zeta_{n-1,\nu})$$

converge uniformly to $p^{*\prime}(z)$ and $p_\nu(z) = \int_{Z_1}^{z} f\nu(t)\,dt$ converge uniformly to $p^*(z)$ for $|z| \leqslant R$, $R > 0$. By Hurwitz' Theorem all the zeros of $p_\nu(z)$ lie interior to $D(0,1)$ and therefore $p_\nu \in \mathcal{P}$ if ν is sufficiently large. However, since Z_1 is a zero of p_ν,

$$I(p_\nu) \geqslant \min\{|Z_1 - \zeta_{k\nu}|, 1 < k < n-1\} > I(Z_1) = I(p^*),$$

contradicting $I(p^*) = I(\mathcal{P})$, the property of extremal polynomials.

Furthermore, Phelps-Rodriguez show that not only one zero but *at least two* zeros of p^* must lie on the unit circle $C(0,1)$. For, if only one such zero, it may be taken without loss of generality as $Z_1 = 1$ with $|Z_j| < 1$ and max $|Z_j| = r$, $j = 2,3,\ldots,n$. We now introduce the polynomial $q(z) = p^*(z+s)$, where $s = (1-r)/2$. The polynomial $q(z)$ has all its zeros $(Z_j - s)$ interior to $D(0,1)$ and its critical points are $(W_k - s)$. Thus $I(q) = I(p^*) = I(\mathcal{P})$, but $q(z)$ cannot be an extremal polynomial since it has no zero on the unit circle. Hence, p^* must have at least two zeros on the unit circle. In fact, from the above reasoning, Phelps-Rodriguez infer that, on the unit circle, every arc of length at least π contains at least one zero Z_j of p^* and, hence, unless $Z_2 = -Z_1$ with $|Z_1| = |Z_2| = 1$, p^* would have *at least three* zeros on the unit circle.

Regarding any zero Z_j of p^* interior to $D(0,1)$, Phelps-Rodriguez establish the inequality

$$[I(Z_j)]^{n-1} \leqslant (1/n)\left(1 + |Z_j|^2\right)\left(1 + |Z_j|\right)^{n-3} \tag{3.1}$$

Thus, if $|Z_j| < 1$, then $I(Z_j) < (2/3)^{1/2}$ for $n = 3$ and $I(Z_j) < 1$ for $n = 4$. Thus for $n = 3$ and 4, Conjecture I is now confirmed. However, if $n \geqslant 5$, inequality (3.1) will not lead to the result $I(Z_j) \leqslant 1$ unless $|Z_j|$ is taken sufficiently small.

4. Critical Point Nearest a Boundary Zero. Whether or not $p(z)$ is extremal in $\mathcal{P}$, let us assume that it has a zero z_1 with $|z_1| = 1$. Let us now show, not merely that $I(z_1) \leqslant 1$ but the sharper result (Fig. 3) that at least one critical point of p lies in $D(z_1/2, 1/2)$. This fact was discovered by Goodman-Rahman-Ratti [**3**]. We give a modified version of the inductive proof as given in Schmeisser [**12**].

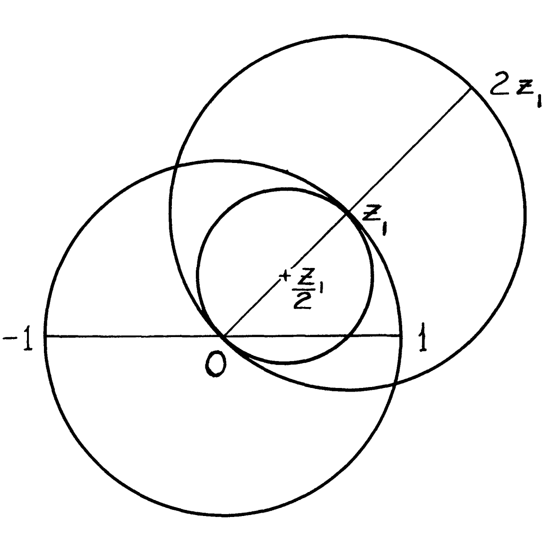


FIG. 3

For $n = 2$, $p(z) = (z - z_1)(z - z_2)$ and $p'(z) = 2(z - \zeta_1)$ with $\zeta_1 = (z_1 + z_2)/2$. In view of $|z_1| = 1$, and $|z_2| \leqslant 1$, $|\zeta_1 - (z_1/2)| \leqslant 1/2$. Let us now assume the theorem to have been proved for $n \leqslant N$; namely that the derivative of $p(z) = (z - z_1) \cdots (z - z_N)$, where $|z_1| = 1$ and $|z_j| \leqslant 1$ for $j \geqslant 2$, has at least one zero in $D(z_1/2, 1/2)$ and consider $p(z) = (z - z_1) \cdots (z - z_{N+1})$ where $|z_1| = 1$ and $|z_j| \leqslant 1$ for $j \geqslant 2$. We introduce the polynomial (polar derivative of p with respect to z_1):

$$q(z) = (N + 1)p(z) - (z - z_1)p'(z) \tag{4.1}$$

which has degree N and a zero at $z = z_1$. If $q(Z) = 0$ but $Z \neq z_1$, we infer from (4.1) that

$$\frac{N+1}{Z - z_1} = \frac{p'(Z)}{p(Z)} = \sum_{j=1}^{N+1} \frac{1}{Z - z_j}.$$

Thus

$$\frac{N}{Z - z_1} = \sum_{j=2}^{N+1} \frac{1}{Z - z_j}. \tag{4.2}$$

If $|Z| > 1$, the function $W = 1/(z - Z)$ maps the disk $D(0, 1)$ conformally upon the closed interior of a circle Γ. The points $w_j = 1/(z_j - Z)$, $j \geqslant 2$, are in the closed interior of Γ but w_1 lies on Γ. According to (4.2),

$$w_1 = (w_2 + w_3 + \cdots + w_{N+1})/N.$$

However, w_1 as a point on Γ cannot be the centroid of points in or on Γ. Thus $q(z)$ has a zero z_1 with $|z_1| = 1$ and all its remaining zeros lie in $D(0, 1)$ so that $q \in \mathcal{P}$ and at least one zero of $q'(z)$ lies in $D(z_1/2, 1/2)$.

Finally, suppose all zeros of $p'(z)$ lie in the circular region $K: |z - (z_1/2)| > 1/2$. By differentiating (4.1), we find

$$q'(z) = Np'(z) - (z - z_1)p''(z). \tag{4.3}$$

Thus, q' has the form,

$$F(z, \lambda) = Nf(z) - (z - \lambda)f'(z),$$

called by Laguerre the "polar derivative" of Nth degree polynomial $f(z)$ with respect to λ. Laguerre proved (Marden [**6**, pp. 49–51]) that, if all the zeros of $f(z)$ lie in a circular region K and if $\lambda \notin K$, then all the zeros of $F(z, \lambda)$ also lie in K. It follows therefore, since $|z_1| = 1$ and hence $z_1 \notin K$, that all the zeros of $q'(z)$ lie in K. Since this contradicts our earlier result that $q'(z)$ has at least one zero in $D(z_1/2, 1/2)$, we infer that not all zeros of $p'(z)$ lie in K and consequently at least one zero of $p'(z)$ lies in $D(z_1/2, 1/2)$, as was to be proved.

Since $D(z_1/2, 1/2) \subset D(z_1, 1)$, we have shown $I(z_1) \leqslant 1$ for any $p \in \mathcal{P}$ for which $p(z_1) = 0, |z_1| = 1$. If we now apply inequality (3.1) to zeros $Z_j, |Z_j| < 1$, we infer that Conjecture I is valid for $n = 2$, 3 and 4. The case $n = 5$ was established by a different method by Meier-Sharma [**8**].

5. Convex Hull of the Zeros. Let us denote by $H(p)$ the convex hull of the zeros of $p(z)$; that is, the smallest convex polygon containing all the zeros of $p(z)$. Since the zeros of $p(z)$ lie in $D(0, 1)$, so does $H(p) \subset D(0, 1)$. (Fig. 4.)

We begin as in Schmeisser [**12**] with the case that all the vertices of $H(p)$ lie on the unit circle $|z| = 1$. If $z_1, z_2, \ldots, z_m (m \leqslant n)$ are at the vertices of $H(p)$ and $z_{m+1}, z_{m+2}, \ldots, z_n$ are interior to $H(p)$, we already know from §4 that at least one critical point ζ_j lies in $D(z_j/2, 1/2)$ for $1 \leqslant j \leqslant m$. Let us now show that

$$H(p) \subset T = \bigcup_{j=1}^{m} D(z_j/2, 1/2).$$

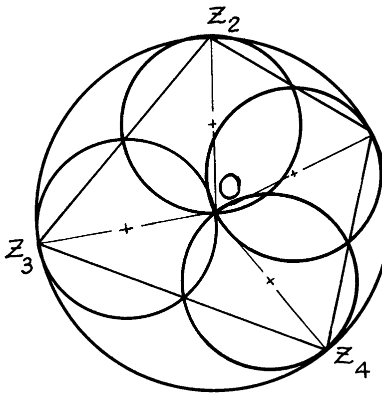


FIG. 4

That is, the region $H(p)$ is completely covered by the set of discs $D(z_j/2, 1/2)$ with $j = 1, 2, \ldots, m$. First, any point z on the side of $H(p)$ joining vertices z_1 and z_2 has the representation $z = \lambda z_1 + (1-\lambda)z_2$ where $0 \leqslant \lambda \leqslant 1$. If point z is closer to z_1 than to z_2, then $(1/2) \leqslant \lambda \leqslant 1$ and so this point lies in $D(z_1/2, 1/2)$ since

$$|z - (z_1/2)| = |[\lambda - (1/2)]z_1 + (1-\lambda)z_2| \leqslant |[\lambda - (1/2)] + (1-\lambda)| = 1/2.$$

Hence the entire side joining z_1 to z_2 lies in T. Similarly all other sides of $H(p)$ and interior points of $H(p)$ lie in T. But, as each disk $D(z_j/2, 1/2)$, $j = 1, 2, \ldots, m$ contains at least one critical point ζ_k of $p(z)$, it follows that Conjecture I is valid for $p(z)$ if its $H(p)$ has all vertices on the unit circle $|z| = 1$.

The case that not all vertices of $H(p)$ lie on the unit circle has been investigated by Schmeisser **[14]** only where $H(p)$ is a triangle ABC with points A and B on circle $|z| = 1$ and C interior to the circle. He also treats the case that at most one vertex of triangle ABC lies on circle $|z| = 1$, by applying a linear transformation $w = az + b$ which reduces this case to that in which two or three vertices of the triangle are on circle $|z| = 1$.

6. Application of Classical Composition Theorems. In attempting to validate Conjecture I, some authors have successfully used certain theorems that are well known in the Analytic Theory of Polynomials. We have in §4 already applied Laguerre's Theorem involving the polar derivative of a polynomial. Another is the following theorem due to Szegö (Marden **[6**, pp. 65–66**]**).

Given the polynomials

$$f(z) = \sum_{k=0}^{m} \binom{m}{k} A_k z^k, \qquad g(z) = \sum_{k=0}^{m} \binom{m}{k} B_k z^k,$$

where $\binom{m}{k} = m(m-1)\cdots(m-k+1)/k!$, *we form the polynomial* $h(z) = \sum_{k=0}^{m}\binom{m}{k} A_k B_k z^k$. *If all the zeros of* $f(z)$ *lie in a circular region* A, *then every zero of* $h(z)$ *has the form* $\gamma = -\alpha\beta$ *where* α *is a suitably chosen point in* A *and* β *is a zero of* $g(z)$.

In Schmeisser **[12]**, Szegö's Theorem is applied to polynomials $p(z)$ which have a zero at the origin. Thus,

$$(6.1) \qquad p(z) = z\sum_{j=0}^{n-1} a_j z^j, \qquad p'(z) = \sum_{k=0}^{n-1} a_k(1+k)z^k.$$

Here the polynomial $p(z)$ may be regarded as $h(z)$, being obtained from $p'(z)$ as the $g(z)$, by

composition with the polynomial $f(z) = F(z)$, where

$$F(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} z^k \Big/ (k+1).$$

But

$$(6.2) \qquad zF(z) = \sum_{k=0}^{n-1} \binom{n-1}{k} \int_0^z t^k\,dt = \int_0^z (1+t)^{n-1}\,dt$$
$$= (1/n)\left[(1+z)^n - 1\right].$$

The zeros of $F(z)$, being $\beta_j = -1 + \exp(2\pi j/n)$; $j = 1, 2, \ldots, n-1$, all lie on the disk $|z+1| \leqslant 1$. Hence, taking $F(z)$ as the $f(z)$ of Szegö's Theorem, we infer that each zero z_j of $p(z)$ has the form $z_j = -\alpha\zeta_k$ where $\alpha \in D(-1,1)$ and $\beta = \zeta_k$ is some zero of $p'(z)$. Thus

$$|z_j - \zeta_k| = |-\alpha\zeta_k - \zeta_k| = |\alpha + 1|\,|\zeta_k| \leqslant 1.$$

This proves that Conjecture I is true for polynomials of form (6.1).

A more elaborate application of Szegö's Theorem is given in Rubinstein [10] in order to verify Conjecture I for polynomials of degree $n = 3$ and 4. Due to the result in §4 for boundary zeros, it suffices to show that for any zeros x_0 of $p(z)$ with $|x_0| < 1$, at least one critical point lies in $D(x_0, 1)$. Without loss of generality we may take x_0 as real with $0 < x_0 < 1$. Writing $p(z) = (z - x_0)q(z)$, we translate the origin to x_0 and introduce the Taylor expansion about $x = x_0$. Thus,

$$f(Z) = p'(Z + x_0) = (d/dZ)\left[Zq(Z + x_0)\right]$$
$$= q(Z + x_0) + Zq'(Z + x_0)$$
$$= \sum_{k=0}^{n-1}\left[q^{(k)}(x_0)/k!\right]Z^k + \sum_{k=0}^{n-2}\left[q^{(k+1)}(x_0)/k!\right]Z^k$$

$$(6.3) \qquad f(Z) = \sum_{k=0}^{n-1}(1+k)\left[q^{(k)}(x_0)/k!\right]Z^k$$

$$(6.4) \qquad g(Z) = \sum_{k=0}^{n-1}\left[\binom{n-1}{k}\Big/(1+k)\right]Z^k$$

$$(6.5) \qquad h(Z) = q(Z + x_0) = \sum_{k=0}^{n-1}\left[q^{(k)}(x_0)/k!\right]Z^k.$$

Let us assume, contrary to Conjecture I, that $p'(z)$ has no zeros within unit distance of the zero x_0; that is, $f(Z) = p'(Z + x_0) \neq 0$ for $[Z] \leqslant 1$. Thus, all the zeros of $f(Z)$ lie in the circular region A: $|Z| > 1$. According to (6.2), the zeros of $g(Z)$ are

$$\beta_k = -1 + \exp(2\pi ki/n), \quad k = 1, 2, \ldots, n-1.$$

Applying Szegö's Theorem, we infer that the zeros Z_j of $h(Z)$ have the form

$$Z_j = -\alpha\beta_k = \alpha\left[1 - \exp(2\pi ki/n)\right]$$

where $|\alpha| > 1$ so that $|Z_j| > 2\sin(\pi/n)$. Since $Zh(Z) = p(Z + x_0)$, the zeros $z_j = Z_j + x_0$ of $p(z)$, other than $z = x_0$, lie in the intersection of the two circular regions of the Z-plane:

$$|Z + x_0| < 1, \qquad |Z| > 2\sin(\pi/n);$$

that is, in the z-plane:

$$(6.6) \qquad |z| < 1, \qquad |z - x_0| > 2\sin(\pi/n).$$

Continuing further, Rubinstein [10] introduces a third plane, the w-plane, and the polynomial

$$R(w) = p(w - 1 + x_0) = (w - 1)\prod_{j=1}^{n-1}\left[w - (1 + z_j - x_0)\right].$$

Hence,

$$R'(w) = p'(w - 1 + x_0) = n\prod_{j=1}^{n-1}\left[w - (1 + \zeta_j - x_0)\right].$$

The assumption that no ζ_j lies in disk $D(x_0, 1)$ is equivalent to saying that no zero of $R'(w)$ lies in the disk $|w - 1| \leqslant 1$. The restrictions (6.6) on the zeros of $R(w)$ other than $w = 1$ now become for $n = 3$ or 4

$$|w - 1 + x_0| < 1, \qquad |w - 1| > 2\sin(\pi/n) \geqslant \sqrt{2}\,. \tag{6.7}$$

Setting $w = u + iv$, we obtain from (6.7) by squaring both sides:

$$u^2 + v^2 - 2(1 - x_0)u + x_0^2 - 2x_0 < 0 \tag{6.8}$$

$$u^2 + v^2 - 2u > 1. \tag{6.9}$$

Multiplying (6.9) by $(1 - x_0)$ and subtracting the product from (6.8), we obtain

$$x_0(u^2 + v^2) < 2x_0 - x_0^2 - 1 + x_0 = 3x_0 - x_0^2 - 1,$$

$$u^2 + v^2 < 3 - \left[x_0 + (1/x_0)\right] < 1 \tag{6.10}$$

due to $x_0 + 1/x_0 > 2$, since $0 < x_0 < 1$.

Thus $R(w)$ is a polynomial which by definition has a zero at $w = 1$ and by (6.10) has all its other zeros in the disk $|w| < 1$. Thus $R(w) \in \mathcal{P}$ and by §4 it has at least one critical point in the disk $|w - 1| < 1$. But, as this contradicts the assumption that $R'(w) \neq 0$ in $|w - 1| \leqslant 1$, it follows that the earlier assumption $f(Z) = p'(Z + x_0) \neq 0$ for $|Z| \leqslant 1$ was wrong and thus $p(z)$ has at least one critical point in disk $D(x_0, 1)$ when $n = 3$ or 4. Thus together with the results of §4, this validates the Conjecture I when $n = 3$ or 4 again.

The Szegö Theorem is also used in Rahman [9] to determine the shortest distance from point $z = a$ to a critical point of the polynomial $[(z - a)p(a)]$ where $p(z)$ has all its zeros in the disk $|z| \leqslant 1$.

As an alternative to using the Szegö Composition Theorem, there is Grace's Theorem from which the Szegö Theorem is derived. Grace's Theorem states the following (Marden [**6**, pp. 60–63]):

Let $f(z) = \sum_{k=0}^{m}\binom{m}{k}A_k z^k$ *and* $g(z) = \sum_{k=0}^{m}\binom{m}{k}B_k z^k$, *be apolar polynomials*; *that is*

$$\sum_{k=0}^{m}(-1)^k\binom{m}{k}A_k B_{m-k} = 0. \tag{6.11}$$

Then any circular region containing all the zeros of f or g contains at least one zero of the other.

There being only one condition (6.11) on the coefficients of f and g, an infinite number of polynomials $g(Z)$ can be found apolar to the polynomial $f(Z)$ given by (6.3). If such a polynomial g can be found that has all its zeros in the disk $|Z| \leqslant 1$, Conjecture I would be established. However, so far in this search the present author has succeeded only in the cases $n = 3$ and 4, already treated by Rubinstein [**10**].

7. Further Conjectures. All the evidence so far produced has been favorable to Conjecture I; no contrary example has as yet been found. The evidence includes, not only the situations discussed in the preceding sections, but the validation of the Conjecture when $n = 5$ (Meier-Sharma [**8**]), and for the lacunary polynomials

$$z^{n_0} + a_1 z^{n_1} + a_2 z^{n_2} + a_3 z^{n_3} \text{ with } n_0 > n_1 > n_2 > n_3$$

and

$$\sum_{k=0}^{m} a_k z^{n_k} \text{ with } n_k \geqslant 3k - 2,\ n_0 < n_1 < \cdots < n_m$$

(Schmeisser [**13**], [**14**]) as well as for real polynomials of the form $p(z) = z^n - a_1 z^{n-1} - \cdots - a_n$ with $a_k \geqslant 0$ for all k (Schmeisser [**13**]).

In the general case, Conjecture I may hinge upon the determination of the extremal polynomials $p^* \in \mathcal{P}$. In §3, it is proved that at least two zeros of $p^*(z)$ must lie on the unit circle $|z| = 1$. Phelps-Rodriguez [**16**] have further conjectured that all the zeros of $p^*(z)$ must lie on the unit circle and in fact be equally spaced on the unit circle. That is, they have proposed the following:

CONJECTURE II. *The extremal polynomials of the family* $\mathcal{P}$ *have the form* $p^*(z) = z^n - \exp(i\alpha), 0 \leqslant \alpha < 2\pi$.

Another conjecture is the following:

CONJECTURE III. *If* $p(z) = \Pi_{j=1}^{n}(z - z_j) \in \mathcal{P}$, *then at least one critical point of* $p(z)$ *lies in each disk* $D(z_j/2, 1 - (|z_j|/2))$, $j = 1, 2, \ldots, n$. (Fig. 5.)

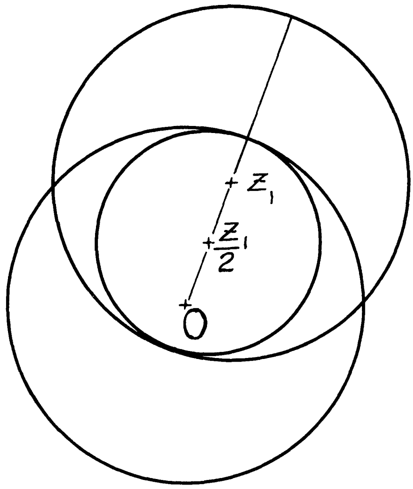


FIG. 5

This conjecture was made both in Goodman-Rahman-Ratti [3] and Schmeisser [**12**]. Conjecture III reduces to the result of §4 when $|z_j| = 1$. If correct, Conjecture III would give a sharper result than Conjecture I in that $D(z_j/2, 1 - (|z_j|/2)) \subset D(z_j, 1)$.

Conjecture III has, however, been proved in Vernon [**15**] for zeros z_j such that $|z_j| < \delta(n)$ where $\delta(n) = 1$ for $n < 4$ but decreases with increasing n, approaching zero as $n \to \infty$. This result is not surprising since Conjecture III is true for $z_j = 0$ (see §6) and since the critical points are continuous functions of the z_j.

As an example illustrating the three conjectures, let us take the polynomial $p(z) = 1 + z + z^2 + z^3$ which has the zeros $z_1 = i$, $z_2 = -i$ and $z_3 = -1$ and for which the derivative $p(z) = 1 + 2z + 3z^2$ has the zeros $\zeta_1 = (-1 + \sqrt{2}\,i)/3$ and $\zeta_2 = (-1 - \sqrt{2}\,i)/3$. A calculation shows that

$$|z_1 - \zeta_1| = |z_2 - \zeta_2| = 0.6249 \quad \text{and} \quad |z_3 - \zeta_1| = |z_3 - \zeta_2| = 0.8165,$$

verifying Conjecture I and indicating that this $p(z)$ is not an extremal polynomial since both

numbers are less than one. A further calculation shows that

$$|\zeta_1 - (z_1/2)| = |\zeta_2 - (z_2/2)| = |-2 + (2\sqrt{2} - 3)i|/6 < 1/2 = 1 - (|z_1|/2) = 1 - (|z_2|/2)$$

$$|\zeta_1 - (z_3/2)| = |\zeta_2 - (z_3/2)| = |1 + \sqrt{2}\,i|/6 < 1/2 = 1 - (|z_3|/2)$$

verifying Conjecture III.

As yet, none of the three conjectures has been fully established. It is clear that, if correct, either Conjecture II or III would imply the correctness of Conjecture I. However, even if Conjecture I were validated independently of Conjectures II and III, it would still be interesting and important to prove or disprove Conjecture II and III.

The author is grateful to his engineering colleague, Professor Emeritus Webster Christman, for the drawings accompanying this paper.

References

1. D. A. Brannan, On a conjecture of Ilieff, Proc. Cambridge Philos. Soc., 64 (1968) 83–85.
2. F. Gacs, On polynomials whose zeros are in the unit disk, J. Math. Anal. Appl., 36 (1971) 627–637.
3. A. W. Goodman, Q. I. Rahman, and J. S. Ratti, On the zeros of a polynomial and its derivative, Proc. Amer. Math. Soc., 21 (1969) 273–274.
4. W. K. Hayman, Research Problems in Function Theory, London, 1967, p. 25, prob. 4.5.
5. Andre Joyal, On the zeros of a polynomial and its derivative, J. Math. Anal. Appl., 26 (1969) 315–317.
6. M. Marden, Geometry of Polynomials, Math. Surveys no. 3, American Mathematical Society, 1966.
7. ______, On the critical points of a polynomial, Tensor (to appear).
8. A. Meir and A. Sharma, On Ilyeff's conjecture, Pacific J. Math., 31 (1969) 459–467.
9. O. I. Rahman, On the zeros of a polynomial and its derivative, Pacific J. Math., 41 (1972) 525–528.
10. Zalman Rubinstein, On a problem of Ilyeff, Pacific J. Math., 26 (1968) 159–161. Also unpublished lecture notes.
11. E. B. Saff and J. B. Twomey, A note on the location of critical points of polynomials, Proc, Amer. Math. Soc., 27 (1971) 303–308.
12. Gerhard Schmeisser, Bemerkungen zu einer Vermutung von Ilieff, Math. Z., 111 (1969) 121–125.
13. G. Schmeisser, Zur lage der kritischen Punkte eines Polynoms, Rend. Sem. Mat. Univ. Padova, 46 (1971) 405–415.
14. ______, On Ilieff's conjecture, Math. Z., 156 (1977) 165–173.
15. S. Vernon, On the critical points of polynomials, Proc. Roy. Irish Acad. Sect. A, 78 (1978) no. 20, 195–198.
16. D. Phelps and R. S. Rodriguez, Some properties of extremal polynomials for the Ilieff Conjecture, Kodai Math. Sem. Rpt. 24 (1972) 172–175.

MISCELLANEA

100.

POETRY

$$(f+g)^n = \sum_{i=0}^{n} \binom{n}{i} f^i g^{n-i}$$

$$(fg)^{(n)} = \sum_{i=0}^{n} \binom{n}{i} f^{(i)} g^{(n-i)}$$

–Ray Bobo, Department of Mathematics, Georgetown University

numbers are less than one. A further calculation shows that

$$|\zeta_1 - (z_1/2)| = |\zeta_2 - (z_2/2)| = |-2 + (2\sqrt{2} - 3)i|/6 < 1/2 = 1 - (|z_1|/2) = 1 - (|z_2|/2)$$

$$|\zeta_1 - (z_3/2)| = |\zeta_2 - (z_3/2)| = |1 + \sqrt{2}\,i|/6 < 1/2 = 1 - (|z_3|/2)$$

verifying Conjecture III.

As yet, none of the three conjectures has been fully established. It is clear that, if correct, either Conjecture II or III would imply the correctness of Conjecture I. However, even if Conjecture I were validated independently of Conjectures II and III, it would still be interesting and important to prove or disprove Conjecture II and III.

The author is grateful to his engineering colleague, Professor Emeritus Webster Christman, for the drawings accompanying this paper.

References

1. D. A. Brannan, On a conjecture of Ilieff, Proc. Cambridge Philos. Soc., 64 (1968) 83–85.
2. F. Gacs, On polynomials whose zeros are in the unit disk, J. Math. Anal. Appl., 36 (1971) 627–637.
3. A. W. Goodman, Q. I. Rahman, and J. S. Ratti, On the zeros of a polynomial and its derivative, Proc. Amer. Math. Soc., 21 (1969) 273–274.
4. W. K. Hayman, Research Problems in Function Theory, London, 1967, p. 25, prob. 4.5.
5. Andre Joyal, On the zeros of a polynomial and its derivative, J. Math. Anal. Appl., 26 (1969) 315–317.
6. M. Marden, Geometry of Polynomials, Math. Surveys no. 3, American Mathematical Society, 1966.
7. ______, On the critical points of a polynomial, Tensor (to appear).
8. A. Meir and A. Sharma, On Ilyeff's conjecture, Pacific J. Math., 31 (1969) 459–467.
9. O. I. Rahman, On the zeros of a polynomial and its derivative, Pacific J. Math., 41 (1972) 525–528.
10. Zalman Rubinstein, On a problem of Ilyeff, Pacific J. Math., 26 (1968) 159–161. Also unpublished lecture notes.
11. E. B. Saff and J. B. Twomey, A note on the location of critical points of polynomials, Proc, Amer. Math. Soc., 27 (1971) 303–308.
12. Gerhard Schmeisser, Bemerkungen zu einer Vermutung von Ilieff, Math. Z., 111 (1969) 121–125.
13. G. Schmeisser, Zur lage der kritischen Punkte eines Polynoms, Rend. Sem. Mat. Univ. Padova, 46 (1971) 405–415.
14. ______, On Ilieff's conjecture, Math. Z., 156 (1977) 165–173.
15. S. Vernon, On the critical points of polynomials, Proc. Roy. Irish Acad. Sect. A, 78 (1978) no. 20, 195–198.
16. D. Phelps and R. S. Rodriguez, Some properties of extremal polynomials for the Ilieff Conjecture, Kodai Math. Sem. Rpt. 24 (1972) 172–175.

MISCELLANEA

100.

POETRY

$$(f+g)^n = \sum_{i=0}^{n} \binom{n}{i} f^i g^{n-i}$$

$$(fg)^{(n)} = \sum_{i=0}^{n} \binom{n}{i} f^{(i)} g^{(n-i)}$$

–Ray Bobo, Department of Mathematics, Georgetown University

THE RIESZ REPRESENTATION THEOREM REVISITED

DONALD G. HARTIG

Mathematics Department, California Polytechnic State University, San Luis Obispo, CA 93407

The purpose of this note is to give a formulation of the celebrated Riesz Representation Theorem (RRT) using the language of categories. Once this has been done we will outline a proof due to D. J. H. Garling [2] that is especially appropriate in this context and deserves to be better known. Garling's proof is mainly topological as opposed to the more standard arguments as found for example in [7]. While the measure theory is reduced to a bare minimum, it is assumed that the student has been exposed to enough measure theory to make the statement of RRT both meaningful and significant. The few topological facts that are required should be accessible to most students after a first course in general topology.

1. Categorical Formulation of RRT. The theorem under consideration states that any bounded linear form $\phi: C(X) \to \mathbb{R}$ can be represented by a unique signed Baire measure μ on X via the equation

$$\phi(f) = \int_X f\,d\mu, \qquad f \in C(X).$$

Here, as always, X is a compact, Hausdorff space and $C(X)$ is the Banach space of continuous, real-valued functions on X. In fact much more is actually true and the exact nature of the relationship between the linear form ϕ and the Baire measure μ is more readily described from the vantage point provided by category theory. Before continuing the reader might profit from a glance at [3].

The category of compact, Hausdorff spaces and continuous maps between them will be denoted by *Comp*. The category of Banach spaces and continuous (i.e., bounded) linear maps between them will be designated by *Ban*. Given an object X in *Comp* there are three objects in *Ban* that are relevant to the present discussion: $C(X)$ described above, its Banach space dual $C(X)^*$ and $M(X)$, the Banach space of signed Baire measures on X. (The σ-algebra of Baire sets in X will be denoted by $\mathfrak{B}_X$; it is the smallest σ-algebra rendering each $f \in C(X)$ a measurable function and is generated by the compact G_δ-sets in X.) Given a morphism $\alpha: X \to Y$ in *Comp* there are also three morphisms in *Ban* that are immediately available:

$$\begin{aligned} \alpha^{\#}: C(Y) &\to C(X) && \text{defined by } \alpha^{\#}(g) = g \circ \alpha \\ \alpha^{\#\#}: C(X)^* &\to C(Y)^* && \text{defined by } \alpha^{\#\#}(\psi) = \psi \circ \alpha^{\#} \\ \alpha^*: M(X) &\to M(Y) && \text{defined by } \alpha^*(\mu) = \mu \circ \alpha^{-1} \end{aligned}$$

At this point it can be easily verified that each of the three maps $\alpha \to \alpha^{\#}$, $\alpha \to \alpha^{\#\#}$ and $\alpha \to \alpha^*$ is functorial. In the earlier note [3] we were interested in the first functor, RRT is a statement about the latter two which we will denote by C^* and M respectively, i.e., $C^*(\alpha) = \alpha^{\#\#}$ and $M(\alpha) = \alpha^*$.

Now given a signed Baire measure μ in $M(X)$ we define $\phi_\mu: C(X) \to \mathbb{R}$ as

$$\phi_\mu(f) = \int_X f\,d\mu.$$

Donald Hartig has a bachelor's degree from Rensselaer Polytechnic Institute and a Master's from the University of Wisconsin-Milwaukee. His doctorate, earned in 1970, is from the University of California at Santa Barbara under the direction of Ky Fan. He has taught at Ohio University, the United States Naval Academy and is currently an Associate Professor at California Polytechnic State University in San Luis Obispo, California. Fluent in Spanish, thanks to his wife who is from Barcelona, Professor Hartig and his family travel often to Spain and enjoy hiking in the Spanish Pyrenees.

The resulting map $\iota_X: M(X) \to C(X)^*$ given by $\iota_X(\mu) = \phi_\mu$ is, to use Mac Lane's terminology [5], "natural in X." That is to say, for any morphism $\alpha: X \to Y$ in *Comp,* the diagram induced in *Ban:*

$$\begin{array}{ccc} C(X)^* & \xrightarrow{\alpha^{\#\#}} & C(Y)^* \\ \iota_X \uparrow & & \uparrow \iota_Y \\ M(X) & \xrightarrow{\alpha^*} & M(Y) \end{array} \qquad \ldots\ldots\ldots (*)$$

is commutative. In the language of category theory the family of maps $\iota = \{\iota_X | X \in Comp\}$ constitutes a **natural transformation** from the functor M to the functor C^*. (To verify that (*) commutes, it must be shown that

$$\int_X g \circ \alpha \, d\mu = \int_Y g \, d\alpha^*(\mu)$$

for each $g \in C(Y)$ and $\mu \in M(X)$. In view of the fact that this equality reduces to the definition of $\alpha^*(\mu)$ when $g = \chi_B$, the characteristic function of a Baire set in Y, the verification is quite straightforward and is left as an exercise.)

Since a measure μ is determined by the integrals it produces and $C(X)$ is dense in each space $L^1(\mu)$ the connecting morphisms ι_X are seen to be one-to-one. With a little more work it can even be shown that each ι_X is norm-preserving and, of course, our initial statement of RRT simply says that ι_X is onto. A natural transformation having the property that each connecting morphism is an isomorphism (one-to-one and onto for the category *Ban*) is called a **natural equivalence.** The full story of RRT can now be stated as the

THEOREM. *The natural transformation $\iota = \{\iota_X\}$ is a natural equivalence from M to C^*.*

2. Proof of RRT. In view of our remarks preceding the statement of the theorem, its proof boils down to showing that each ι_X is an onto map. The key idea is to observe that this is quite easy to do for spaces X that are very fragmented. For example, the argument is trivial when X is discrete (hence finite) and we will see that it is only slightly more difficult for the so-called "extremally disconnected" spaces. It turns out that such spaces are plentiful enough in *Comp* to allow us to easily extend the proof to all ι_X's via the diagram displayed in (*). The only result we will need that is of a purely measure theoretic nature is Caratheodory's Extension Theorem as found, for example, in Chapter 12 of [6].

A topological space is called **extremally disconnected** when each open set has an open closure. This is equivalent to requiring that each pair of disjoint open sets have disjoint closures. In such spaces the algebra $\mathcal{C}$ of subsets that are open and closed (the so-called **clopen** sets) is very large and plays a decisive role in both topological and measure-theoretic considerations. This rather broad generalization is made precise in the following lemma.

LEMMA 1. *Let X be extremally disconnected in Comp. Then the Baire sets in X are generated by the algebra $\mathcal{C}$ of clopen sets and the simple functions based on clopen sets are uniformly dense in $C(X)$.*

Proof. Let $\mathcal{S}$ denote the σ-algebra generated by $\mathcal{C}$. Since each set in $\mathcal{C}$ is a compact G_δ, $\mathcal{S} \subset \mathcal{B}_X$. To show that $\mathcal{B}_X \subset \mathcal{S}$ it will suffice to establish that each $f \in C(X)$ is an $\mathcal{S}$-measurable function. Given $f \in C(X)$ and real number c let $A = \{x \in X | f(x) \leqslant c\}$. For $n = 1, 2, \ldots$ define $A_n = \{x \in X | f(x) < c + 1/n\}$. Because A_n is open, its closure $\bar{A}_n$ is clopen and from

$$A_n \subset \bar{A}_n \subset \left\{ x \in X \middle| f(x) \leqslant c + \frac{1}{n} \right\}$$

it follows that $A = \cap_n \bar{A}_n$ implying that $A \in \mathcal{S}$.

To prove the second assertion choose $f \in C(X)$ and let n be a fixed positive integer. For $k = 0, \pm 1, \pm 2, \dots$ let

$$A_k = \left\{ x \in X \middle| \frac{k}{n} < f(x) < \frac{k+1}{n} \right\}$$

As above, the closure B_k of each set A_k is a clopen set and since at most finitely many of the B_k's are nonempty, the sets $B = \cup_k B_k$ and $X \setminus B$ are also clopen. It follows easily that the sets

$$C_k = \left\{ x \in X \setminus B | f(x) = \frac{k}{n} \right\}, \quad k = 0, \pm 1, \pm 2, \dots$$

are clopen and clearly the simple function

$$g = \sum_k \frac{k}{n} [\chi_{B_k} + \chi_{C_k}]$$

is within $1/n$ of f uniformly on X. End of proof.

Lemma 1 provides a basis for the proof of RRT for the subcategory of *Comp* consisting of the extremally disconnected spaces. Let X be such a space and let ϕ be a nonnegative linear form on $C(X)$, that is $\phi(f) \geqslant 0$ whenever $f \geqslant 0$. We immediately obtain a nonnegative finitely additive set function μ on the algebra $\mathcal{C}$ of clopen sets by defining $\mu(A) = \phi(\chi_A)$. This measure is even countably additive on $\mathcal{C}$ since compactness implies that any countable family of disjoint clopen sets whose union is also clopen can have at most a finite number of nonempty members. Caratheodory's Extension Theorem, together with Lemma 1, allows us to extend μ to a Baire measure on X. The second part of Lemma 1 shows that this extension of μ represents the linear form ϕ. Since an arbitrary member of $C(X)^*$ can be obtained as the difference of nonnegative forms, this establishes our assertion.

The extension to arbitrary X in *Comp* will be accomplished after we state two topological lemmas. Lemma 2 says that extremally disconnected compact Hausdorff spaces abound while Lemma 3 shows that they appear throughout the category *Comp.*

LEMMA 2. *The Stone-Čech compactification of a discrete space is extremally disconnected.*

LEMMA 3. *Any compact Hausdorff space is the continuous image of an extremally disconnected compact Hausdorff space.*

The proof of Lemma 2 is left as an exercise, Lemma 3 follows directly from Lemma 2.

Proof of RRT. Let Y be compact and Hausdorff. Choose an onto morphism $\alpha : X \to Y$ in *Comp* where X is extremally disconnected. One easily sees that $\alpha^{\#}$ is a norm isomorphism and an application of the Hahn-Banach Theorem will show that $\alpha^{\#\#}$ is also onto. We already know that ι_X is onto and (referring to the commutative diagram (*)) it follows that ι_Y is onto.

3. Related Results. To my knowledge V. S. Varadarajan first had the idea of shifting the main burden of the proof of RRT to topological considerations. His approach is from a slightly different angle and the interested reader will find additional information in his paper **[9]**. A brief version of Varadarajan's proof can also be found in **[1]**. Garling's ingenious arrangement of the proof of RRT can also be found in the recent text on functional analysis **[4]** by R. B. Holmes. This book is warmly recommended for its wealth of interesting applications and insightful proofs. The categorical formulation of RRT is well known. It is one more example of the usefulness of this point of view and is described in Semadeni's book **[8]**, see Proposition 18.4.3.

References

1. L. Baez-Duarte, $C(X)^*$ and Kolmogorov's Consistency Theorem for Cantor spaces, Studies in Appl. Math., 49 (1970) 401–403.

2. D. J. H. Garling, A "short" proof of the Riesz representation theorem, Proc. Cambridge Philos. Soc., 73 (1973) 459–460.

3. D. G. Hartig, An important functor in analysis and topology, this MONTHLY, 85 (1978) 41–43.

4. R. B. Holmes, Geometric Functional Analysis and its Applications, Springer-Verlag, New York, Heidelberg, Berlin, 1975.

5. S. Mac Lane, Categories for the Working Mathematician, Springer-Verlag, New York, Heidelberg, Berlin, 1971.

6. H. Royden, Real Analysis, 2nd ed., Macmillan, New York, 1968.

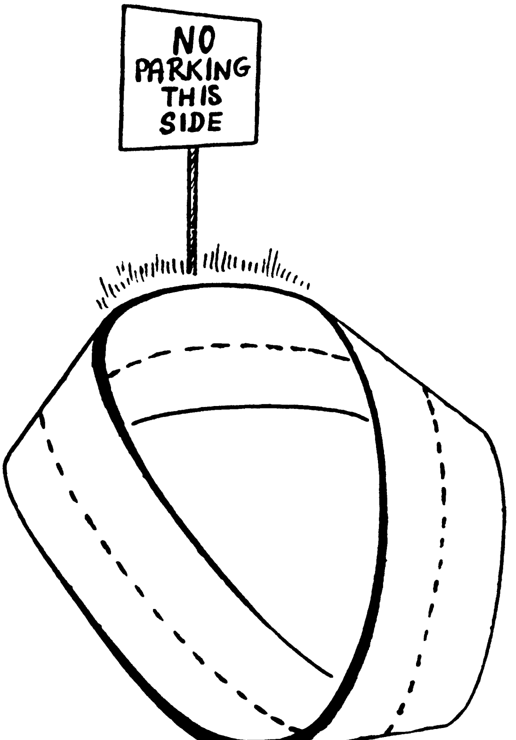
7. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.

8. Z. Semadeni, Banach Spaces of Continuous Functions, vol. 1, Polish Scientific Publishers, Warsaw, 1971.

9. V. S. Varadarajan, On a theorem of F. Riesz concerning the form of linear functionals, Fund. Math., 46 (1958) 209–220.

MISCELLANEA

101.



— James R. Martino
Inspired by the difficulties
of parking in time for class.

2. D. J. H. Garling, A "short" proof of the Riesz representation theorem, Proc. Cambridge Philos. Soc., 73 (1973) 459–460.

3. D. G. Hartig, An important functor in analysis and topology, this MONTHLY, 85 (1978) 41–43.

4. R. B. Holmes, Geometric Functional Analysis and its Applications, Springer-Verlag, New York, Heidelberg, Berlin, 1975.

5. S. Mac Lane, Categories for the Working Mathematician, Springer-Verlag, New York, Heidelberg, Berlin, 1971.

6. H. Royden, Real Analysis, 2nd ed., Macmillan, New York, 1968.

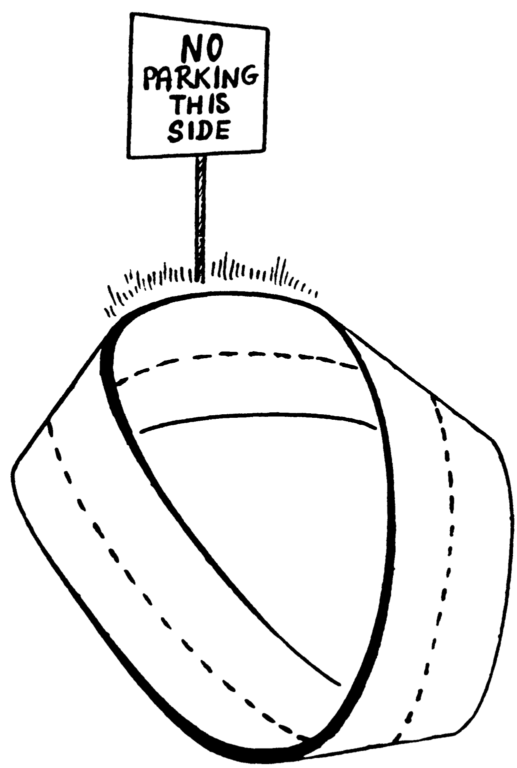
7. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.

8. Z. Semadeni, Banach Spaces of Continuous Functions, vol. 1, Polish Scientific Publishers, Warsaw, 1971.

9. V. S. Varadarajan, On a theorem of F. Riesz concerning the form of linear functionals, Fund. Math., 46 (1958) 209–220.

MISCELLANEA

101.



— James R. Martino
Inspired by the difficulties
of parking in time for class.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

FUNCTIONS WITH A PROPER LOCAL MAXIMUM IN EACH INTERVAL

E. E. POSEY AND J. E. VAUGHAN
Department of Mathematics, University of North Carolina at Greensboro, Greensboro, NC 27412

Let $f\colon R \to R$ be a real function. We say that f has a *proper local maximum* at $x \in R$ provided there exists an open neighborhood V of x such that $f(y) < f(x)$ for all $y \in V - \{x\}$. In 1900 A. Schoenflies [3, p. 158] noted that for any function f, the set of points at which f has a proper local maximum is at most countable. (This is easy to see: Let B denote the set of all open intervals in R with rational end points; so B is a countable base for the usual topology on R. For each proper local maximum x of f, pick $V_x \in B$ such that $f(y) < f(x)$ for all $y \in V_x - \{x\}$. The correspondence $x \mapsto V_x$ is one-to-one.)

It is trivial to sketch the graph of a continuous function $f\colon R \to R$ which has a proper local maximum at infinitely many points in [0, 1], say at each point in the set $\{1/n\colon n = 1, 2, \cdots\} \cup \{0\}$. This brings up the question: Does there exist a continuous function $f\colon R \to R$ which has a proper local maximum at each point of a countable dense set? It is not easy to visualize the graph of such a continuous function, but the answer to the question is known to be "yes." In [**1**, p. 63, Thm. 3] it was proved that each parameter function of a completely wild arc in R^3 will be a continuous function having a proper local maximum at each point of a countable dense set. This result provides an "existence theorem" for such functions, but the concept of such arcs is not very intuitive.

The purpose of this note is to construct, in a straightforward way, a simple example of a continuous function $f\colon R \to R$ which has a proper local maximum at each rational point. This function will be the uniform limit of functions which are easy to visualize. In this sense the construction is analogous to several well-known constructions of continuous nowhere differentiable functions (e.g. [**2**, p. 141, Thm. 7.18]), and provides a reasonable exercise for a real analysis course.

EXAMPLE: A function $f\colon R \to R$ with a proper local maximum at each point of a set everywhere dense in R.

Let $\{r_n\}_{n=1}^{\infty}$ be a set (of distinct points) that is everywhere dense in the real numbers, and let

$$g(x) = 1 - \mathrm{Min}[1, |x|],$$

and let

$$f(x) = \sum_{n=1}^{\infty} g_n(x)$$

where $g_n(x) = A_n g((x - r_n)/w_n)$, and the positive real numbers A_n and w_n will be specified.

The definition of g_n will be by induction on n. It is clear that $g_n(x) = 0$ for $|x - r_n| > w_n$, and we denote the interval $|x - r_n| \leqslant w_n$ by I_n and call it the interval associated with g_n. We also define

$$f_n(x) = \sum_{j=1}^{n} g_j(x).$$

In general we require (1) $0 < A_n \leqslant 2^{-n}$; (2) the end points of I_n be distinct from all points r_j, $j = 1,2,3\ldots$; and (3) if $j < n$, then either (3′) $r_n \notin I_j$ and w_n is chosen so that I_n and I_j are disjoint or (3″) $r_n \in I_j$ and w_n is chosen so that I_n is strictly interior to that half of I_j in which r_n lies, e.g.,

$$\text{if } r_j - w_j < r_n < r_j,$$

then

$$r_j - w_j < r_n - w_n < r_n + w_n < r_j.$$

Choose by induction positive constants A_n and w_n so that (1), (2), and (3) are satisfied. It is clear that because of (1) f is well defined and continuous regardless of how the constants w_n are specified. In the following paragraphs we will change some of the constants A_n and w_n; we do so with the understanding that conditions (1), (2), and (3) are still to be satisfied. The conditions (1), (2), and (3) on the choice of A_n and w_n will not be repeated at each further stage of the construction.

Let A_1 and w_1 be chosen as above, thus defining f_1. As a sum of "triangular" functions g_j, f_n will be constructed to satisfy:

(a) f_n has a proper local maximum at r_n, being linear and increasing on $(r_n - w_n, r_n)$ and linear and decreasing on $(r_n, r_n + w_n)$.

(b) If $I_n \subsetneqq I_j$, then $j < n$ and $f_n(r_n) < f_j(r_j)$.

Suppose f_{n-1} has been defined and satisfies (a) and (b). If I_n is exterior to $I_j, j = 1,2,\ldots, n-1$, then A_n and w_n are not further modified and (a) and (b) are valid for f_n. Otherwise let h be the maximum integer not exceeding $n-1$ such that $r_n \in I_h$. Then by the restrictions imposed previously I_n is strictly interior to $(r_h - w_h, r_h)$ or to $(r_h, r_h + w_h)$, and f_h is linear on I_n. By (a) $f_h(r_n - w_n)$ and $f_h(r_n + w_n)$ are both less than $f_h(r_h)$. Hence A_n can be taken sufficiently small so that $f_n(r_n) < f_h(r_h)$. If $r_h \in I_j$ for $j < h$, then $I_h \subset I_j$, and thus by (b) $f_h(r_h) < f_j(r_j)$ so $f_n(r_n) < f_j(r_j)$ and (b) holds. With A_n fixed we may now decrease w_n until (a) is satisfied. Indeed, the maximum and minimum slopes of g_j are $\pm A_j/w_j$ and thus it suffices to require

$$A_n/w_n > \sum_{j=1}^{n-1} A_j/w_j$$

to insure that (a) holds. Thus the induction is complete and conditions (1), (2), (3), and (a), (b), hold generally.

The function f so defined has a proper local maximum at each point r_n. Indeed, let $x \in I_n$, $x \neq r_n$, and we shall prove that $f(x) < f(r_n)$. If $x \notin I_j$ for $j = n+1, n+2,\ldots$ then $f(x) = f_n(x) < f_n(r_n) \leqslant f(r_n)$ by (a). Suppose now that h is the smallest index exceeding n such that $x \in I_h$. If a largest index k exceeding n for which $x \in I_k$ exists, we have

$$f(x) = f_k(x) < f_k(r_k) < f_n(r_n) \leqslant f(r_n)$$

by (a) and (b). If no such largest k exists, then $f(x) = \lim_{i\to\infty} f_{k_i}(x)$ where $x \in I_{k_i}$, $i = 1,2,3,\ldots$ and all $k_i > h$. Then

$$f_{k_i}(x) < f_{k_i}(r_{k_i}) < f_h(r_h) < f_n(r_n)$$

and hence $f(x) \leqslant f_h(r_h) < f_n(r_n) \leqslant f(r_n)$, and the proof is complete.

References

1. E. E. Posey, Proteus forms of wild and tame arcs, Duke Math. J., 31 (1964) 63–72.

2. Walter Rudin, Principles of Mathematical Analysis, McGraw-Hill, New York, 1964.

3. A. Schoenflies, Die Entwickelung der Lehre von den Punktmannigfaltigkeiten, Bericht, erstattet der Deutschen Mathematiker-Vereinigung, 1900.

A SPACE FILLING CURVE

LIU WEN
Department of Mathematics, Graduate Center of the City University of New York, New York, NY 10036
Hebei Institute of Technology, Tianjin, the People's Republic of China

The purpose of this paper is to give an example of a space filling curve which is simpler than previous examples (cf. [**1**], 455–458).

Divide the interval $[0, 1]$ into ten equal subintervals:

$$\delta_k = [k/10, (k+1)/10], \quad k = 0, 1, \ldots, 9.$$

Let f and g be continuous real valued functions defined on $[0, 1]$ which satisfy the following conditions:

(i)
$$f(t) = \begin{cases} 0 & \text{when} \quad t \in \delta_1 \cup \delta_3, \\ 1 & \text{when} \quad t \in \delta_5 \cup \delta_7; \end{cases}$$

$$g(t) = \begin{cases} 0 & \text{when} \quad t \in \delta_1 \cup \delta_5, \\ 1 & \text{when} \quad t \in \delta_3 \cup \delta_7; \end{cases}$$

(ii)
$$f(0) = f(1) \quad \text{and} \quad g(0) = g(1).$$

Extending f and g to $(-\infty, +\infty)$ periodically, we get two continuous periodic functions which are still denoted f and g respectively.

Let

$$\phi(t) = \sum_{k=1}^{\infty} \frac{1}{2^k} f(10^{k-1}t); \quad \psi(t) = \sum_{k=1}^{\infty} \frac{1}{2^k} g(10^{k-1}t).$$

Obviously ϕ and ψ are continuous. We will now easily show that

$$\begin{cases} x = \phi(t) \\ y = \psi(t) \end{cases} \qquad 0 \leqslant t \leqslant 1$$

is a curve filling the unit square.

In fact, let x and y be arbitrary real numbers in $(0, 1)$ with base 2 binary expansions

$$x = .x_1 x_2 \ldots x_k \ldots ; \quad \text{each } x_k = 0 \text{ or } 1; \quad y = .y_1 y_2 \ldots y_k \ldots ; \quad \text{each } y_k = 0 \text{ or } 1.$$

Let

$$t_k = \begin{cases} 1 & \text{if} \quad x_k = 0 \quad \text{and} \quad y_k = 0 \\ 3 & \text{if} \quad x_k = 0 \quad \text{and} \quad y_k = 1 \\ 5 & \text{if} \quad x_k = 1 \quad \text{and} \quad y_k = 0 \\ 7 & \text{if} \quad x_k = 1 \quad \text{and} \quad y_k = 1. \end{cases}$$

Now let t be the number whose base 10 decimal expansion is

$$t = .t_1 t_2 \ldots t_k \ldots .$$

It follows immediately from the definitions that

$$x = \phi(t), \quad y = \psi(t),$$

and so the curve passes through every point of the unit square.

Reference

1. E. W. Hobson, Theory of Functions of a Real Variable, vol. 1, Dover, New York, 1957.

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

SUMMING POWER SERIES WITH POLYNOMIAL COEFFICIENTS

JOHN KLIPPERT
Department of Mathematics and Computer Science, James Madison University, Harrisonburg, Virginia 22807

Third semester calculus students quickly learn that

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{for} \quad |x| < 1$$

by showing that the sequence of partial sums

$$S_k = \sum_{n=0}^{k} x^n = \frac{1-x^{k+1}}{1-x}, \quad x \neq 1.$$

But by rewriting a power series in terms of finite differences, we can almost as easily sum *any* power series with polynomial coefficients. Consider the series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ with positive radius of convergence. Then

$$f(x) = a_0 + x\phi(x) \quad \text{where} \quad \phi(x) = \sum_{n=1}^{\infty} a_n x^{n-1} = \sum_{n=0}^{\infty} a_{n+1} x^n.$$

So

$$\begin{aligned}(1-x)\phi(x) &= \sum_{n=0}^{\infty} a_{n+1}x^n - x\left(\sum_{n=1}^{\infty} a_n x^{n-1}\right)\\ &= \sum_{n=0}^{\infty} a_{n+1}x^n - \sum_{n=1}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_{n+1}x^n - \sum_{n=0}^{\infty} a_n x^n + a_0\\ &= a_0 + \sum_{n=0}^{\infty} (a_{n+1} - a_n)x^n = a_0 + \sum_{n=0}^{\infty} \Delta a_n x^n,\end{aligned}$$

with $\Delta a_n = a_{n+1} - a_n, \quad n = 0, 1, 2, \ldots .$

Therefore if $x \neq 1$ and $x \in D$, the domain of f, we have

$$\phi(x) = \frac{a_0}{1-x} + \frac{1}{1-x} \sum_{n=0}^{\infty} \Delta a_n x^n.$$

Thus

$$f(x) = a_0 + x\phi(x) = a_0 + \frac{a_0 x}{1-x} + \frac{x}{1-x} \sum_{n=0}^{\infty} \Delta a_n x^n.$$

That is,

$$\sum_{n=0}^{\infty} a_n x^n = \frac{a_0}{1-x} + \frac{x}{1-x} \sum_{n=0}^{\infty} \Delta a_n x^n, \quad x \neq 1, \quad x \in D. \tag{1}$$

Putting $\Delta^2 a_n = \Delta(\Delta a_n) = \Delta a_{n+1} - \Delta a_n$ (and inductively,

$$\Delta^p a_n = \Delta\left(\Delta^{p-1}a_n\right) = \Delta^{p-1}a_{n+1} - \Delta^{p-1}a_n, \quad p \in N)$$

and applying (1) to $\sum_{n=0}^{\infty}\Delta a_n x^n$, we have

$$\sum_{n=0}^{\infty} \Delta a_n x^n = \frac{\Delta a_0}{1-x} + \frac{x}{1-x}\sum_{n=0}^{\infty}\Delta^2 a_n x^n$$

and so

$$\sum_{n=0}^{\infty} a_n x^n = \frac{a_0}{1-x} + \frac{x}{1-x}\left(\frac{\Delta a_0}{1-x} + \frac{x}{1-x}\sum_{n=0}^{\infty}\Delta^2 a_n x^n\right)$$

$$= \frac{a_0}{1-x} + \frac{x\Delta a_0}{(1-x)^2} + \left(\frac{x}{1-x}\right)^2\sum_{n=0}^{\infty}\Delta^2 a_n x^n.$$

Repeating p times in succession, we get

$$\sum_{n=0}^{\infty} a_n x^n = \frac{a_0}{1-x} + \frac{x\Delta a_0}{(1-x)^2} + \cdots + \frac{x^{p-1}\Delta^{p-1}a_0}{(1-x)^p} + \left(\frac{x}{1-x}\right)^p\sum_{n=0}^{\infty}\Delta^p a_n x^n.$$

That is,

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{k=0}^{p-1}\Delta^k a_0\frac{x^k}{(1-x)^{k+1}} + \left(\frac{x}{1-x}\right)^p\sum_{n=0}^{\infty}\Delta^p a_n x^n, \quad x \neq 1, x \in D \tag{2}$$

where we define $\Delta^0 a_0 = a_0$.

Now if $a_n = P(n)$, a polynomial of degree $\leqslant p-1$, we have

$$\Delta^p P(n) = 0, \quad n = 0, 1, 2, \ldots.$$

Observing that the radius of convergence of $\sum_{n=0}^{\infty}P(n)x^n$ is 1, we have according to (2) that

$$\sum_{n=0}^{\infty} P(n)x^n = \sum_{k=0}^{p-1}\Delta^k P(0)\frac{x^k}{(1-x)^{k+1}}, \quad |x| < 1. \tag{3}$$

For example, if $P(n) = n^2 + n + 1$, then $\Delta a_n = 2n + 1$, $\Delta^2 a_n = 2$ and $\Delta^3 a_n = 0$. In particular, $P(0) = 1$, $\Delta P(0) = 2$, $\Delta^2 P(0) = 2$. Putting $p = 3$ into (3), we have

$$\sum_{n=0}^{\infty} P(n)x^n = \sum_{k=0}^{2}\Delta^k P(0)\frac{x^k}{(1-x)^{k+1}}$$

and so

$$\sum_{n=0}^{\infty}(n^2+n+1)x^n = P(0)\frac{1}{1-x} + \Delta P(0)\frac{x}{(1-x)^2} + \Delta^2 P(0)\frac{x^2}{(1-x)^3}$$

$$= \frac{1}{1-x} + \frac{2x}{(1-x)^2} + \frac{2x^2}{(1-x)^3}.$$

Finally, we remark that the transform (2) can also be used to accelerate the convergence of a series when $\Delta^p a_n$ converges to 0 faster than a_n (assuming that a_n *does* converge to 0).

Reference

1. Konrad Knopp, Theory and Application of Infinite Series, Blackie and Son Limited, London, 1966, pp. 230–273.

E 2993. *Proposed by Michael Larsen, Student, Harvard University.*

Let $\alpha_1, \alpha_2, \dots, \alpha_n$ be complex numbers such that $\Sigma_1^n \alpha_i^m$ is an integer for every positive integer m. Show that the polynomial $\Pi_1^n(x - \alpha_i)$ has integer coefficients.

E 2994. *Proposed by John E. Wetzel, University of Illinois.*

Let Z^d be the set of lattice points in $\mathbb{R}^d$, the set of points all of whose coordinates are integers. For any subset A of Z^d write $\mathcal{D}(A)$ for the distance set of A, the set of distances XY for X, Y in A.

(a) Show that Z^d can be partitioned into two subsets A and B so that neither $\mathcal{D}(A)$ nor $\mathcal{D}(B)$ equals $\mathcal{D}(Z^d)$.

(b) Show that if $A \cup B = Z^d$ but $A \cap B \neq \varnothing$, then either $\mathcal{D}(A)$ or $\mathcal{D}(B)$ must equal $\mathcal{D}(Z^d)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Fair Division of a Cake

E 2895 [1981, 444]. *Proposed by M. Vulis, CUNY Graduate Center.*

The children at a birthday party sit around a table and divide a homogeneous cake according to the following rules: Initially the birthday child divides the cake into two equal portions, giving one portion to one of the two neighboring children and keeping the other. At the next and succeeding steps, some pair of adjacent children combine their portions and divide equally. (Thus if two adjacent children have portions a and b before division and share them, each has a portion $\frac{1}{2}(a + b)$ after the division. Note that either a or b can be zero.)

The process terminates in a finite number of steps with a fair division of the cake. How many children are at the party?

Solution by David M. Wells, Pennsylvania State University, New Kensington. There are 4 children at the party. To establish this, identify the children with the group of integers (mod n). Let $f_j(k)$ denote the amount of cake held by child k after j steps. The proof depends on two propositions, each established by induction on j.

PROPOSITION 1: *There are integers k_1 and k_2 for which $f_j(k_1) \leqslant f_j(k_1 + 1) \leqslant \cdots \leqslant f_j(k_2)$ and $f_j(k_2) \geqslant f_j(k_2 + 1) \geqslant \cdots \geqslant f_j(k_1)$.*

PROPOSITION 2: *If $n > 4$ there are at most two consecutive values of k for which $f_j(k) = 1/n$.*

The proof of the first is tedious but straightforward. To prove the second, suppose the statement is true after j steps. We may assume with no loss of generality that $f_j(1) = 1/n$, $f_j(2) > 1/n$, and $f_j(3) < 1/n$. Proposition 1 then guarantees that $f_j(k) \leqslant 1/n$ for $3 \leqslant k \leqslant n$, and by assumption $f_j(k) < 1/n$ for $3 \leqslant k \leqslant n - 1$. Thus $f_{j+1}(2) = f_{j+1}(3) > 1/n$, since $\Sigma f_{j+1}(k) = 1$.

Proposition 2 insures that $n \leqslant 4$, and the statement of the problem implies that $n \geqslant 3$. However, $f_j(k)$ is always a rational number whose denominator is a power of 2, so $n \neq 3$.

Also solved by O. P. Lossers (Netherlands), K. Wayland, and the proposer.

An Eulerian Circuit with No Crossings

E 2897 [1981, 537]. *Proposed by David Singmaster, Polytechnic of the South Bank, London, UK.*

Consider a planar graph G and a path P in it. If the path visits a vertex two or more times, we say there is a crossing at the vertex if the path would cross itself when the vertex is viewed as a road intersection; e.g., in the figure on page 288, a path containing $\dots,1,7,4,\dots,2,7,5,\dots$ has a crossing, while a path containing $\dots,1,7,2,\dots,3,7,6,\dots,4,7,5,\dots$ has no crossing. Prove or

E 2993. *Proposed by Michael Larsen, Student, Harvard University.*

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be complex numbers such that $\Sigma_1^n \alpha_i^m$ is an integer for every positive integer m. Show that the polynomial $\Pi_1^n(x - \alpha_i)$ has integer coefficients.

E 2994. *Proposed by John E. Wetzel, University of Illinois.*

Let Z^d be the set of lattice points in $\mathbb{R}^d$, the set of points all of whose coordinates are integers. For any subset A of Z^d write $\mathfrak{D}(A)$ for the distance set of A, the set of distances XY for X, Y in A.

(a) Show that Z^d can be partitioned into two subsets A and B so that neither $\mathfrak{D}(A)$ nor $\mathfrak{D}(B)$ equals $\mathfrak{D}(Z^d)$.

(b) Show that if $A \cup B = Z^d$ but $A \cap B \neq \varnothing$, then either $\mathfrak{D}(A)$ or $\mathfrak{D}(B)$ must equal $\mathfrak{D}(Z^d)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Fair Division of a Cake

E 2895 [1981, 444]. *Proposed by M. Vulis, CUNY Graduate Center.*

The children at a birthday party sit around a table and divide a homogeneous cake according to the following rules: Initially the birthday child divides the cake into two equal portions, giving one portion to one of the two neighboring children and keeping the other. At the next and succeeding steps, some pair of adjacent children combine their portions and divide equally. (Thus if two adjacent children have portions a and b before division and share them, each has a portion $\frac{1}{2}(a + b)$ after the division. Note that either a or b can be zero.)

The process terminates in a finite number of steps with a fair division of the cake. How many children are at the party?

Solution by David M. Wells, Pennsylvania State University, New Kensington. There are 4 children at the party. To establish this, identify the children with the group of integers (mod n). Let $f_j(k)$ denote the amount of cake held by child k after j steps. The proof depends on two propositions, each established by induction on j.

PROPOSITION 1: *There are integers k_1 and k_2 for which $f_j(k_1) \leqslant f_j(k_1 + 1) \leqslant \cdots \leqslant f_j(k_2)$ and $f_j(k_2) \geqslant f_j(k_2 + 1) \geqslant \cdots \geqslant f_j(k_1)$.*

PROPOSITION 2: *If $n > 4$ there are at most two consecutive values of k for which $f_j(k) = 1/n$.*

The proof of the first is tedious but straightforward. To prove the second, suppose the statement is true after j steps. We may assume with no loss of generality that $f_j(1) = 1/n$, $f_j(2) > 1/n$, and $f_j(3) < 1/n$. Proposition 1 then guarantees that $f_j(k) \leqslant 1/n$ for $3 \leqslant k \leqslant n$, and by assumption $f_j(k) < 1/n$ for $3 \leqslant k \leqslant n - 1$. Thus $f_{j+1}(2) = f_{j+1}(3) > 1/n$, since $\Sigma f_{j+1}(k) = 1$.

Proposition 2 insures that $n \leqslant 4$, and the statement of the problem implies that $n \geqslant 3$. However, $f_j(k)$ is always a rational number whose denominator is a power of 2, so $n \neq 3$.

Also solved by O. P. Lossers (Netherlands), K. Wayland, and the proposer.

An Eulerian Circuit with No Crossings

E 2897 [1981, 537]. *Proposed by David Singmaster, Polytechnic of the South Bank, London, UK.*

Consider a planar graph G and a path P in it. If the path visits a vertex two or more times, we say there is a crossing at the vertex if the path would cross itself when the vertex is viewed as a road intersection; e.g., in the figure on page 288, a path containing $\ldots, 1, 7, 4, \ldots, 2, 7, 5, \ldots$ has a crossing, while a path containing $\ldots, 1, 7, 2, \ldots, 3, 7, 6, \ldots, 4, 7, 5, \ldots$ has no crossing. Prove or

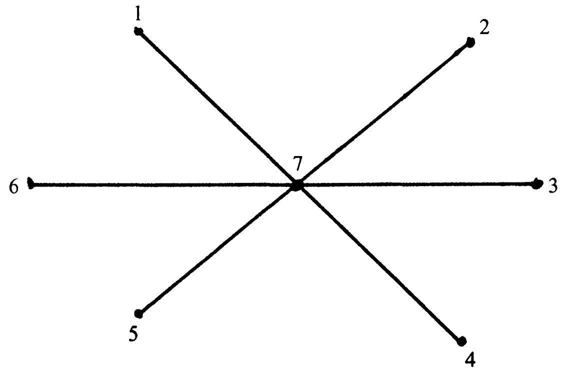


FIG. 1

disprove: A connected planar graph with every vertex of even degree has an Eulerian circuit with no crossings.

Solution by Jerrold W. Grossman, Oakland University, and E. M. Reingold, University of Illinois. This problem is solved in the affirmative in "Construction of Planar Eulerian Multigraphs," in *Proc. Tenth Southeastern Conf. on Combinatorics, Graph Theory and Computing*, 1979, pp. 123–130, by J. Abraham & A. Kotzig. Here is a simpler proof, by induction on the maximum degree d of a vertex and, within the class of graphs with the same maximum degree, the number of vertices with the maximum degree. Clearly if $d = 2$, then every Euler circuit has no crossings. Assume the inductive hypothesis and let v be a vertex of the graph G of degree d. Split v into two vertices, one with degree two and incident to two neighboring edges of G, the other incident to all its other edges of G. If the resulting graph is connected, then its Euler circuit with no crossings clearly induces such a circuit in G. If not, then the Euler circuits without crossings in the two components can be spliced together at v to form the desired Euler circuit in G. See the subjoined figure, in which the dotted sectors contain no edges.

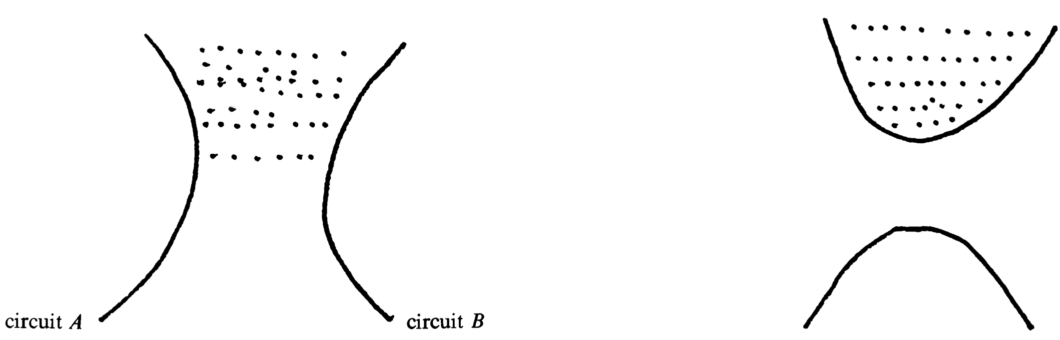


FIG. 2

A related and I believe still unsolved problem is due to H. Fleischner: if the plane, Eulerian graph G is 3-connected (i.e., removal of two vertices cannot disconnect the graph) and has no loops or multiple edges, then is there necessarily an Euler circuit which at each vertex leaves along one of the two edges neighboring the edge from which it entered?

Also solved by R. B. Eggleton & D. K. Skilton (Australia), M. L. Marx, E. Scheinerman, R. Statman, and the proposer.

Marx cited his article (with L. Lovasz), *A forbidden substructure characterization of Gauss codes*, Szeged Acta, 38 (1976) 115–119.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by August 31, 1983. The solver's full post-office address should be on each sheet.

6423. *Proposed by P. R. Halmos, Indiana University.*

If A is a bounded linear operator on a Hilbert space, and if

$$\|1 - A\| < 1,$$

then A is invertible. That assertion, not only for operators on Hilbert spaces, but for arbitrary elements of Banach algebras, is usually proved by consideration of the infinite series $\Sigma_{n=0}^{\infty}(1 - A)^n$, and, therefore, the proof uses completeness. Is the assertion true without completeness? Is there, in other words, a bounded linear operator A on an inner product space, such that $\|1 - A\| < 1$ but A is not invertible?

6424. *Proposed by Michael J. Dixon, California State University, Chico.*

Let m and k be positive integers. Show that

$$S(m, k) = \sum_{n=1}^{\infty} \frac{1}{n(mn + k)}$$

is rational if and only if $k \equiv 0 \pmod{m}$.

6425. *Proposed by J. Miles and L. Rubel, University of Illinois.*

For the Koebe function $K(z) = z/(1 - z)^2$, show that there exists a bounded analytic function $B(z)$ in the unit disc $\mathbb{D}$ such that $F(z) = K(z) + B(z)$ maps $\mathbb{D}$ onto the whole complex plane $\mathbb{C}$. Is the same still true for an arbitrary analytic function $K(z)$ in $\mathbb{D}$ that maps $\mathbb{D}$ onto a dense subset of $\mathbb{C}$?

6426. *Proposed by J. L. Brenner, Palo Alto, California.*

(a) Show that every finite group can be embedded in one or another of the finite groups $\mathrm{PSL}(n, q)$. (b) The same for $O^+(n, q)$. (c)* For fixed q, find the smallest $n = n(k)$ such that the symmetric group S_k can be embedded in $\mathrm{PSL}(n, q)$. (Take representative values of q, k.)

SOLUTIONS OF ADVANCED PROBLEMS

The Hankel Matrix Associated with a Power Series

6313 [1980, 675]. *Proposed by L. W. Tu, University of Michigan.*

Let $f(t) = \Sigma a_i t^i$ be a monic power series ($a_0 = 1$; $a_i = 0$ if $i < 0$). Let $\Delta_{pq}(f)$ be the $q \times q$ Hankel matrix with (i, j) element a_{p-i+j}. Prove that $\det \Delta_{pq}(1/f) = (-1)^{pq} \det \Delta_{qp}(f)$.

[Note that, with $q = 1$, this gives a determinantal formula for the coefficients of $1/f$. In the special case that $f(t)$ is the quotient of polynomials of respective degrees p, q, the result seems to be known.]

I. *Solution by J. G. Horne, University of Georgia.* Associate with any power series $f(t) = \Sigma a_i t^i$ the infinite matrix $A(f)$ whose (i, j) entry is a_{j-i}, with the agreement that $a_k = 0$ if $k < 0$. This association is an isomorphism from the ring of such power series under convolution to the ring of infinite upper triangular matrices. Now assume $a_0 = 1$ and let $1/f(t) = \Sigma b_i t^i$. Let r, s be any two integers with $r \geqslant s \geqslant 1$. Let A, B be the $r \times r$ "upper left hand corner" submatrices of $A(f)$ and $B(1/f)$ respectively. Then $\det A = \det B = 1$ and the relation $f \cdot 1/f = 1$ guarantees that $B =$

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by August 31, 1983. The solver's full post-office address should be on each sheet.

6423. *Proposed by P. R. Halmos, Indiana University.*

If A is a bounded linear operator on a Hilbert space, and if

$$\|1 - A\| < 1,$$

then A is invertible. That assertion, not only for operators on Hilbert spaces, but for arbitrary elements of Banach algebras, is usually proved by consideration of the infinite series $\Sigma_{n=0}^{\infty}(1 - A)^n$, and, therefore, the proof uses completeness. Is the assertion true without completeness? Is there, in other words, a bounded linear operator A on an inner product space, such that $\|1 - A\| < 1$ but A is not invertible?

6424. *Proposed by Michael J. Dixon, California State University, Chico.*

Let m and k be positive integers. Show that

$$S(m,k) = \sum_{n=1}^{\infty} \frac{1}{n(mn+k)}$$

is rational if and only if $k \equiv 0 \pmod{m}$.

6425. *Proposed by J. Miles and L. Rubel, University of Illinois.*

For the Koebe function $K(z) = z/(1 - z)^2$, show that there exists a bounded analytic function $B(z)$ in the unit disc $\mathbb{D}$ such that $F(z) = K(z) + B(z)$ maps $\mathbb{D}$ onto the whole complex plane $\mathbb{C}$. Is the same still true for an arbitrary analytic function $K(z)$ in $\mathbb{D}$ that maps $\mathbb{D}$ onto a dense subset of $\mathbb{C}$?

6426. *Proposed by J. L. Brenner, Palo Alto, California.*

(a) Show that every finite group can be embedded in one or another of the finite groups $\mathrm{PSL}(n, q)$. (b) The same for $O^+(n, q)$. (c)* For fixed q, find the smallest $n = n(k)$ such that the symmetric group S_k can be embedded in $\mathrm{PSL}(n, q)$. (Take representative values of q, k.)

SOLUTIONS OF ADVANCED PROBLEMS

The Hankel Matrix Associated with a Power Series

6313 [1980, 675]. *Proposed by L. W. Tu, University of Michigan.*

Let $f(t) = \Sigma a_i t^i$ be a monic power series ($a_0 = 1$; $a_i = 0$ if $i < 0$). Let $\Delta_{pq}(f)$ be the $q \times q$ Hankel matrix with (i, j) element a_{p-i+j}. Prove that $\det \Delta_{pq}(1/f) = (-1)^{pq} \det \Delta_{qp}(f)$.

[Note that, with $q = 1$, this gives a determinantal formula for the coefficients of $1/f$. In the special case that $f(t)$ is the quotient of polynomials of respective degrees p, q, the result seems to be known.]

I. *Solution by J. G. Horne, University of Georgia.* Associate with any power series $f(t) = \Sigma a_i t^i$ the infinite matrix $A(f)$ whose (i, j) entry is a_{j-i}, with the agreement that $a_k = 0$ if $k < 0$. This association is an isomorphism from the ring of such power series under convolution to the ring of infinite upper triangular matrices. Now assume $a_0 = 1$ and let $1/f(t) = \Sigma b_i t^i$. Let r, s be any two integers with $r \geqslant s \geqslant 1$. Let A, B be the $r \times r$ "upper left hand corner" submatrices of $A(f)$ and $B(1/f)$ respectively. Then $\det A = \det B = 1$ and the relation $f \cdot 1/f = 1$ guarantees that $B =$

A^{-1}. If k is any positive integer, $n = \binom{r}{k}^2$ and C is any square matrix, let $C^{(k)}$ denote the $n \times n$ matrix consisting of the $k \times k$ minors of C, arranged lexicographically. Let $\mathrm{Adj}^{(k)}C$ denote the transpose of the matrix obtained by replacing each entry of $C^{(k)}$ by its complementary cofactor. Now $(A^{(k)})^{-1} = (A^{-1})^{(k)}$, and when $\det A = 1$, $(A^{(k)})^{-1} = \mathrm{Adj}^{(k)}A$. Therefore,

$$B^{(s)} = (A^{-1})^{(s)} = (A^{(s)})^{-1} = \mathrm{Adj}^{(s)}A.$$

This implies the following set of identities whenever $j_1, \ldots, j_s, j'_1, \ldots, j'_{r-s}$ are complementary subsets of the indices $\{1, \ldots, r\}$ with $1 \leqslant j_1 < \cdots < j_s \leqslant r$ and $1 \leqslant j'_1 < \cdots < j'_{r-s} \leqslant r$:

$$\det A_1 = (-1)^v \det B_1 \tag{1}$$

where

$$A_1 = \begin{pmatrix} a'_{j_1-1} a'_{j_2-1} & \cdots & a'_{j_{r-s}-1} \\ \vdots & & \vdots \\ a'_{j_1-(r-s)} & \cdots & a'_{j_{r-s}-(r-s)} \end{pmatrix},$$

$$B_1 = \begin{pmatrix} b_{r-s-(j_1-1)} & b_{r-s-(j_1-2)} \ldots & b_{r-j_1} \\ b_{r-s-(j_2-1)} & \cdots & b_{r-j_2} \\ \vdots & & \vdots \\ b_{r-s-(j_s-1)} & \cdots & b_{r-j_s} \end{pmatrix}$$

and $v = \dfrac{(r-s)(r-s+1)}{2} + j'_1 + \cdots + j'_{r-s}$.

The equality (1) follows from the observation that

$$A_1 = A\begin{pmatrix} 1 & 2 \cdots r-s \\ j'_1 & j'_2 \cdots j'_{r-s} \end{pmatrix}$$

and

$$B_1 = B\begin{pmatrix} j_1 & j_2 \cdots j_s \\ r-s+1 & \ldots r \end{pmatrix}$$

and the fact that $B^{(s)} = \mathrm{Adj}^{(s)A.}$

To see that these identities include those in the proposed problem, let p, q, be arbitrary positive integers. Set $r = p + q$, $s = p$. Take $j_i = i$, $i = 1, \ldots, p$, $j'_i = p + i$, $i = 1, \ldots, q$, then note that A_1 reduces to $\Delta_{pq}(f)$, and B_1 reduces to $\Delta_{qp}(1/f)$ and $(-1)^{pq} = (-1)^v$. The set of identities (1) is much larger than this set.

II. *Solution by D. M. Jackson, University of Waterloo, Canada.* We set $f^{-1} = 1 + \Sigma_1^\infty b_i t^i$, and let $B[A]$ be $(p+q) \times (p+q)$ Hankel matrices with constant diagonals

$$0, \ldots, 0, b_0, b_1, \ldots, b_{p+q-1}[0, \ldots, 0, a_0, a_1, \ldots];$$

$b_0\,[a_0]$ is the principal diagonal. ($b_0 = a_0 = 1$.) Assume without loss of generality that $p > q$, and partition $B[A]$ into a 3×3 matrix of boxes $B_{ij}[A_{ij}]$, of respective principal sizes $q \times q$, $(p-q) \times (p-q)$, $q \times q$. Then since $BA = I$, we have

$$B_{11}A_{12} + B_{12}A_{22} + B_{13}A_{32} = 0, \text{ etc.} \tag{*}$$

We set

$$C_{p\times p} = [B_{11}, B_{12}; B_{21}, B_{22}],$$

$$D_{p\times p} = [A_{12}; A_{13}; A_{22}, A_{23}].$$

In this notation, we have to prove $\det D = (-1)^{pq}\det B_{13}$. (Note $\det C = 1$ since $B_{21} = 0$.) Using (*) and similar relations, and using also $A_{32} = 0$, we find

$$CD = [0, -B_{13}A_{33};\, I, -B_{23}A_{33}].$$

But $\det A_{33} = 1$, and elementary transformations show that

$$\det CD = \left[(-1)^q\right]^{p-q}\det(-B_{13}A_{33}).$$

To see this, we computed $[0, I_{q\times q};\, I_{p-q\times p-q}, 0]CD$. Since $-B_{13}A_{33}$ is a $q\times q$ matrix, this gives

$$\det CD = (-1)^{q(p-q)+q}\det B_{13}.$$

Thus $\det CD = (-1)^{pq}\det B_{13}$, from which the result follows.

The proposer notes that the formula appears to have a long history. He refers to formula (36) in J. Hadamard's *Essai sur l'étude des fonctions données par leur développement de Taylor*, J. Math., 8 (1892) 101–186.

Collections of Topologies

6357 [1981, 623]. *Proposed by Zachary Robinson, student, Massachusetts Institute of Technology.*

Let X be a nonempty set. Let $\mathcal{S}$ be a collection of topologies on X, well ordered by strict inclusion. Let τ be the topology generated by the base $\cup\mathcal{S}$. Prove or disprove: If τ is second countable, then $\mathcal{S}$ is countable.

Solution by Douglas E. Cameron, The University of Akron. The assertion is false. Let X be the set of real numbers $\mathbb{R}$, and let $\ll$ be a well ordering of $\mathbb{R}$. For $\alpha\in\mathbb{R}$, let τ_α be the least topology on $\mathbb{R}$ such that $\mathbb{R}-\{\beta\}\in\tau_\alpha$ for $\alpha\ll\beta$. Then $\cup_{\alpha\in\mathbb{R}}\tau_\alpha$ is a base for $\mathcal{F}$, the topology of finite complements on $\mathbb{R}$. Also $\tau_\alpha\subset\tau_\beta$ if $\alpha\ll\beta$ and $\alpha\neq\beta$. Let τ be the usual topology on $\mathbb{R}$. Then

$$\mathcal{S} = \{\tau\}\cup\{\tau_\alpha \mid \alpha\in\mathbb{R}\}$$

is well ordered by strict set inclusion, $\cup\mathcal{S} = \tau$ is second countable and $|\mathcal{S}| = |\mathbb{R}|$.

Also solved by F. S. Cater, A. K. Desai (India), K. P. Hart (The Netherlands), and the proposer.

A Dirichlet Character Sum

6359. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let χ be a nontrivial Dirichlet character, and let α be complex. Show that $\Sigma\chi(p)(\log p)^\alpha p^{-s}$ converges for $\mathrm{Re}(s)\geq 1$, where the sum is over all primes p.

Solution by the proposer. Let $A(x) = \Sigma_{p\leq x}\chi(p)\log p$ for $x>1$. After grouping the primes $p\leq x$ into congruence classes, the prime number theorem for arithmetic progressions with error term can be applied to show that $A(x) = O(x/(\log x)^B)$ for any constant $B>0$. Thus

$$\begin{aligned}\Sigma\chi(p)(\log p)^\alpha p^{-s} &= \int_{1.5}^{\infty}(\log x)^{\alpha-1}x^{-s}\,dA(x)\\ &= \int_{1.5}^{\infty}A(x)x^{-s-1}\{s(\log x)^{\alpha-1}-(\alpha-1)(\log x)^{\alpha-2}\}\,dx,\end{aligned}$$

and the last integral converges absolutely for $\mathrm{Re}(s)\geq 1$. (The above equality, of course, arises after letting a finite upper index of summation approach infinity.)

Conjugates in Division Rings

6361 [1981, 623]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let D be a division ring, $a\in D^*$, the group of nonzero elements. Suppose that a has only

In this notation, we have to prove $\det D = (-1)^{pq}\det B_{13}$. (Note $\det C = 1$ since $B_{21} = 0$.) Using (*) and similar relations, and using also $A_{32} = 0$, we find

$$CD = [0, -B_{13}A_{33}; I, -B_{23}A_{33}].$$

But $\det A_{33} = 1$, and elementary transformations show that

$$\det CD = \left[(-1)^q\right]^{p-q}\det(-B_{13}A_{33}).$$

To see this, we computed $[0, I_{q\times q}; I_{p-q\times p-q}, 0]CD$. Since $-B_{13}A_{33}$ is a $q \times q$ matrix, this gives

$$\det CD = (-1)^{q(p-q)+q}\det B_{13}.$$

Thus $\det CD = (-1)^{pq}\det B_{13}$, from which the result follows.

The proposer notes that the formula appears to have a long history. He refers to formula (36) in J. Hadamard's *Essai sur l'étude des fonctions données par leur développement de Taylor*, J. Math., 8 (1892) 101–186.

Collections of Topologies

6357 [1981, 623]. *Proposed by Zachary Robinson, student, Massachusetts Institute of Technology.*

Let X be a nonempty set. Let $\mathcal{S}$ be a collection of topologies on X, well ordered by strict inclusion. Let τ be the topology generated by the base $\cup\mathcal{S}$. Prove or disprove: If τ is second countable, then $\mathcal{S}$ is countable.

Solution by Douglas E. Cameron, The University of Akron. The assertion is false. Let X be the set of real numbers $\mathbb{R}$, and let $\ll$ be a well ordering of $\mathbb{R}$. For $\alpha \in \mathbb{R}$, let τ_α be the least topology on $\mathbb{R}$ such that $\mathbb{R} - \{\beta\} \in \tau_\alpha$ for $\alpha \ll \beta$. Then $\cup_{\alpha\in\mathbb{R}}\tau_\alpha$ is a base for $\mathcal{F}$, the topology of finite complements on $\mathbb{R}$. Also $\tau_\alpha \subset \tau_\beta$ if $\alpha \ll \beta$ and $\alpha \neq \beta$. Let τ be the usual topology on $\mathbb{R}$. Then

$$\mathcal{S} = \{\tau\} \cup \{\tau_\alpha | \alpha \in \mathbb{R}\}$$

is well ordered by strict set inclusion, $\cup\mathcal{S} = \tau$ is second countable and $|\mathcal{S}| = |\mathbb{R}|$.

Also solved by F. S. Cater, A. K. Desai (India), K. P. Hart (The Netherlands), and the proposer.

A Dirichlet Character Sum

6359. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let χ be a nontrivial Dirichlet character, and let α be complex. Show that $\Sigma\chi(p)(\log p)^\alpha p^{-s}$ converges for $\mathrm{Re}(s) \geqslant 1$, where the sum is over all primes p.

Solution by the proposer. Let $A(x) = \Sigma_{p\leqslant x}\chi(p)\log p$ for $x > 1$. After grouping the primes $p \leqslant x$ into congruence classes, the prime number theorem for arithmetic progressions with error term can be applied to show that $A(x) = O(x/(\log x)^B)$ for any constant $B > 0$. Thus

$$\begin{aligned}\Sigma\chi(p)(\log p)^\alpha p^{-s} &= \int_{1.5}^{\infty}(\log x)^{\alpha-1}x^{-s}dA(x)\\ &= \int_{1.5}^{\infty}A(x)x^{-s-1}\{s(\log x)^{\alpha-1} - (\alpha-1)(\log x)^{\alpha-2}\}\,dx,\end{aligned}$$

and the last integral converges absolutely for $\mathrm{Re}(s) \geqslant 1$. (The above equality, of course, arises after letting a finite upper index of summation approach infinity.)

Conjugates in Division Rings

6361 [1981, 623]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let D be a division ring, $a \in D^*$, the group of nonzero elements. Suppose that a has only

In this notation, we have to prove $\det D = (-1)^{pq}\det B_{13}$. (Note $\det C = 1$ since $B_{21} = 0$.) Using (*) and similar relations, and using also $A_{32} = 0$, we find

$$CD = [0, -B_{13}A_{33}; I, -B_{23}A_{33}].$$

But $\det A_{33} = 1$, and elementary transformations show that

$$\det CD = \left[(-1)^q\right]^{p-q}\det(-B_{13}A_{33}).$$

To see this, we computed $[0, I_{q\times q}; I_{p-q\times p-q}, 0]CD$. Since $-B_{13}A_{33}$ is a $q \times q$ matrix, this gives

$$\det CD = (-1)^{q(p-q)+q}\det B_{13}.$$

Thus $\det CD = (-1)^{pq}\det B_{13}$, from which the result follows.

The proposer notes that the formula appears to have a long history. He refers to formula (36) in J. Hadamard's *Essai sur l'étude des fonctions données par leur développement de Taylor*, J. Math., 8 (1892) 101–186.

Collections of Topologies

6357 [1981, 623]. *Proposed by Zachary Robinson, student, Massachusetts Institute of Technology.*

Let X be a nonempty set. Let $\mathbb{S}$ be a collection of topologies on X, well ordered by strict inclusion. Let τ be the topology generated by the base $\cup\mathbb{S}$. Prove or disprove: If τ is second countable, then $\mathbb{S}$ is countable.

Solution by Douglas E. Cameron, The University of Akron. The assertion is false. Let X be the set of real numbers $\mathbb{R}$, and let $\ll$ be a well ordering of $\mathbb{R}$. For $\alpha \in \mathbb{R}$, let τ_α be the least topology on $\mathbb{R}$ such that $\mathbb{R} - \{\beta\} \in \tau_\alpha$ for $\alpha \ll \beta$. Then $\cup_{\alpha\in\mathbb{R}}\tau_\alpha$ is a base for $\mathfrak{F}$, the topology of finite complements on $\mathbb{R}$. Also $\tau_\alpha \subset \tau_\beta$ if $\alpha \ll \beta$ and $\alpha \neq \beta$. Let τ be the usual topology on $\mathbb{R}$. Then

$$\mathbb{S} = \{\tau\} \cup \{\tau_\alpha | \alpha \in \mathbb{R}\}$$

is well ordered by strict set inclusion, $\cup\mathbb{S} = \tau$ is second countable and $|\mathbb{S}| = |\mathbb{R}|$.

Also solved by F. S. Cater, A. K. Desai (India), K. P. Hart (The Netherlands), and the proposer.

A Dirichlet Character Sum

6359. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let χ be a nontrivial Dirichlet character, and let α be complex. Show that $\Sigma\chi(p)(\log p)^\alpha p^{-s}$ converges for $\mathrm{Re}(s) \geqslant 1$, where the sum is over all primes p.

Solution by the proposer. Let $A(x) = \Sigma_{p\leqslant x}\chi(p)\log p$ for $x > 1$. After grouping the primes $p \leqslant x$ into congruence classes, the prime number theorem for arithmetic progressions with error term can be applied to show that $A(x) = O(x/(\log x)^B)$ for any constant $B > 0$. Thus

$$\begin{aligned}\Sigma\chi(p)(\log p)^\alpha p^{-s} &= \int_{1.5}^{\infty}(\log x)^{\alpha-1}x^{-s}\,dA(x)\\ &= \int_{1.5}^{\infty}A(x)x^{-s-1}\{s(\log x)^{\alpha-1} - (\alpha-1)(\log x)^{\alpha-2}\}\,dx,\end{aligned}$$

and the last integral converges absolutely for $\mathrm{Re}(s) \geqslant 1$. (The above equality, of course, arises after letting a finite upper index of summation approach infinity.)

Conjugates in Division Rings

6361 [1981, 623]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let D be a division ring, $a \in D^*$, the group of nonzero elements. Suppose that a has only

In this notation, we have to prove $\det D = (-1)^{pq}\det B_{13}$. (Note $\det C = 1$ since $B_{21} = 0$.) Using (*) and similar relations, and using also $A_{32} = 0$, we find

$$CD = [0, -B_{13}A_{33}; I, -B_{23}A_{33}].$$

But $\det A_{33} = 1$, and elementary transformations show that

$$\det CD = \left[(-1)^q\right]^{p-q}\det(-B_{13}A_{33}).$$

To see this, we computed $[0, I_{q\times q}; I_{p-q\times p-q}, 0]CD$. Since $-B_{13}A_{33}$ is a $q \times q$ matrix, this gives

$$\det CD = (-1)^{q(p-q)+q}\det B_{13}.$$

Thus $\det CD = (-1)^{pq}\det B_{13}$, from which the result follows.

The proposer notes that the formula appears to have a long history. He refers to formula (36) in J. Hadamard's *Essai sur l'étude des fonctions données par leur développement de Taylor*, J. Math., 8 (1892) 101–186.

Collections of Topologies

6357 [1981, 623]. *Proposed by Zachary Robinson, student, Massachusetts Institute of Technology.*

Let X be a nonempty set. Let $\mathbb{S}$ be a collection of topologies on X, well ordered by strict inclusion. Let τ be the topology generated by the base $\cup\mathbb{S}$. Prove or disprove: If τ is second countable, then $\mathbb{S}$ is countable.

Solution by Douglas E. Cameron, The University of Akron. The assertion is false. Let X be the set of real numbers $\mathbb{R}$, and let $\ll$ be a well ordering of $\mathbb{R}$. For $\alpha \in \mathbb{R}$, let τ_α be the least topology on $\mathbb{R}$ such that $\mathbb{R} - \{\beta\} \in \tau_\alpha$ for $\alpha \ll \beta$. Then $\cup_{\alpha\in\mathbb{R}}\tau_\alpha$ is a base for $\mathfrak{F}$, the topology of finite complements on $\mathbb{R}$. Also $\tau_\alpha \subset \tau_\beta$ if $\alpha \ll \beta$ and $\alpha \neq \beta$. Let τ be the usual topology on $\mathbb{R}$. Then

$$\mathbb{S} = \{\tau\} \cup \{\tau_\alpha | \alpha \in \mathbb{R}\}$$

is well ordered by strict set inclusion, $\cup\mathbb{S} = \tau$ is second countable and $|\mathbb{S}| = |\mathbb{R}|$.

Also solved by F. S. Cater, A. K. Desai (India), K. P. Hart (The Netherlands), and the proposer.

A Dirichlet Character Sum

6359. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let χ be a nontrivial Dirichlet character, and let α be complex. Show that $\Sigma\chi(p)(\log p)^\alpha p^{-s}$ converges for $\mathrm{Re}(s) \geqslant 1$, where the sum is over all primes p.

Solution by the proposer. Let $A(x) = \Sigma_{p\leqslant x}\chi(p)\log p$ for $x > 1$. After grouping the primes $p \leqslant x$ into congruence classes, the prime number theorem for arithmetic progressions with error term can be applied to show that $A(x) = O(x/(\log x)^B)$ for any constant $B > 0$. Thus

$$\Sigma\chi(p)(\log p)^\alpha p^{-s} = \int_{1.5}^{\infty}(\log x)^{\alpha-1}x^{-s}dA(x)$$
$$= \int_{1.5}^{\infty}A(x)x^{-s-1}\{s(\log x)^{\alpha-1} - (\alpha-1)(\log x)^{\alpha-2}\}\,dx,$$

and the last integral converges absolutely for $\mathrm{Re}(s) \geqslant 1$. (The above equality, of course, arises after letting a finite upper index of summation approach infinity.)

Conjugates in Division Rings

6361 [1981, 623]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let D be a division ring, $a \in D^*$, the group of nonzero elements. Suppose that a has only

finitely many conjugates. Prove that a is in k, the center of D. (Application: a polynomial p with coefficients in k, irreducible over k but with a root in D, has infinitely many roots in D.)

Solution by Fergus J. Gaines, University College, Belfield, Dublin. Since, by Wedderburn's theorem, a finite division ring is a field, we may assume that D is infinite. Let $a \in D^*$ have finitely many conjugates. Let $C(a) = \{x \in D^* : ax = xa\}$. Then $C(a)$ is a subgroup of D^* and its index, which equals the number of conjugates of a in D^*, is finite. Suppose there exists $x \in D^*$ with $x^{-1}ax \neq a$. We claim that $C(a)$ is finite. To see this, note that there are only finitely many elements of the form $(b + x)^{-1}a(b + x)$ where $b \in C(a)$. Now let $b, c \in C(a)$ with

$$(b + x)^{-1}a(b + x) = (c + x)^{-1}a(c + x).$$

Then $(b + x)(c + x)^{-1} = d \in C(a)$. This gives $d = 1$, since $d \neq 1$ implies $x \in C(a)$, because $(d - 1)x = b - dc$. Thus $b = c$. So $C(a)$ is finite and has finite index in D^*. Thus D is finite. This contradiction implies $a \in k$, the centre of D.

Also solved by Jeffrey Bergen, Jerzy Browkin (Poland), F. S. Cater, Thomas C. Craven, Hugh M. Edgar, Lorraine L. Foster, Gary R. Greenfield, A. A. Jagers (The Netherlands), Michael Josephy (Costa Rica), Charles Lanski, Susan Montgomery, R. W. K. Odoni, A. J. Pianzola (Canada), David R. Richman, Tom Sadler, Rony Teitler (U.K.), Ernst Trost (Switzerland), Z. Z. Uoiea, William C. Waterhouse, Gregory P. Wene, and the proposer.

A number of the solvers observed that a solution appears in I. N. Herstein, Conjugates in division rings, *Proc. Amer. Math. Soc.*, 7 (1956) 1021–1022. The solution above is simpler and more elementary than Herstein's.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Geometry, Particles and Fields. By Bjørn Felsäger. Odense University Press, Odense, Denmark, 1982. xx + 643 pp., Dan. Kr. 110.00.

CLIFFORD HENRY TAUBES
Department of Physics, Harvard University, Cambridge, MA 02138

To write a report for a mathematics journal on a manuscript partly about physics is at best a dangerous task. The reason is that, in the century past, mathematics and physics became distinct. It is, therefore, useful to explain first what physics is and does. The distinction between mathematics and physics is this: while physics explores our natural world, mathematics explores all possible worlds. Physics itself is a symbiotic pair. Half of it is experimental physics, which seeks to uncover and describe new phenomena: to investigate, map, probe, and quantify the events that occur in our natural world. It is a glorious crusade, "going where no man has gone before." Perhaps it is our destiny. To the experimental physicist, "data" is the operative word.

The other partner in the symbiosis is the discipline called Theoretical Physics. A theoretical physicist is the modern theologian. He seeks, in almost a biblical sense, to understand the data, to perceive the patterns in the chaos, and ultimately to predict. The operative word is "understand." This, too, is our destiny. (In antiquity, the symbiotic split was less pronounced—the same priests who nightly recorded the positions of the sun, moon, planets and stars also made predictions of eclipses from their data. Now we live in a society of specialists.)

The symbiosis spins as a cycle; the experimentalist discovers new data; the theorist must then

---

Supported in part by Harvard Society of Fellows and the National Science Foundation under Grant PHY79-16812.

finitely many conjugates. Prove that a is in k, the center of D. (Application: a polynomial p with coefficients in k, irreducible over k but with a root in D, has infinitely many roots in D.)

Solution by Fergus J. Gaines, University College, Belfield, Dublin. Since, by Wedderburn's theorem, a finite division ring is a field, we may assume that D is infinite. Let $a \in D^*$ have finitely many conjugates. Let $C(a) = \{x \in D^* : ax = xa\}$. Then $C(a)$ is a subgroup of D^* and its index, which equals the number of conjugates of a in D^*, is finite. Suppose there exists $x \in D^*$ with $x^{-1}ax \neq a$. We claim that $C(a)$ is finite. To see this, note that there are only finitely many elements of the form $(b+x)^{-1}a(b+x)$ where $b \in C(a)$. Now let $b, c \in C(a)$ with

$$(b+x)^{-1}a(b+x) = (c+x)^{-1}a(c+x).$$

Then $(b+x)(c+x)^{-1} = d \in C(a)$. This gives $d = 1$, since $d \neq 1$ implies $x \in C(a)$, because $(d-1)x = b - dc$. Thus $b = c$. So $C(a)$ is finite and has finite index in D^*. Thus D is finite. This contradiction implies $a \in k$, the centre of D.

Also solved by Jeffrey Bergen, Jerzy Browkin (Poland), F. S. Cater, Thomas C. Craven, Hugh M. Edgar, Lorraine L. Foster, Gary R. Greenfield, A. A. Jagers (The Netherlands), Michael Josephy (Costa Rica), Charles Lanski, Susan Montgomery, R. W. K. Odoni, A. J. Pianzola (Canada), David R. Richman, Tom Sadler, Rony Teitler (U.K.), Ernst Trost (Switzerland), Z. Z. Uoiea, William C. Waterhouse, Gregory P. Wene, and the proposer.

A number of the solvers observed that a solution appears in I. N. Herstein, Conjugates in division rings, *Proc. Amer. Math. Soc.*, 7 (1956) 1021–1022. The solution above is simpler and more elementary than Herstein's.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Geometry, Particles and Fields. By Bjørn Felsäger. Odense University Press, Odense, Denmark, 1982. xx + 643 pp., Dan. Kr. 110.00.

CLIFFORD HENRY TAUBES
Department of Physics, Harvard University, Cambridge, MA 02138

To write a report for a mathematics journal on a manuscript partly about physics is at best a dangerous task. The reason is that, in the century past, mathematics and physics became distinct. It is, therefore, useful to explain first what physics is and does. The distinction between mathematics and physics is this: while physics explores our natural world, mathematics explores all possible worlds. Physics itself is a symbiotic pair. Half of it is experimental physics, which seeks to uncover and describe new phenomena: to investigate, map, probe, and quantify the events that occur in our natural world. It is a glorious crusade, "going where no man has gone before." Perhaps it is our destiny. To the experimental physicist, "data" is the operative word.

The other partner in the symbiosis is the discipline called Theoretical Physics. A theoretical physicist is the modern theologian. He seeks, in almost a biblical sense, to understand the data, to perceive the patterns in the chaos, and ultimately to predict. The operative word is "understand." This, too, is our destiny. (In antiquity, the symbiotic split was less pronounced—the same priests who nightly recorded the positions of the sun, moon, planets and stars also made predictions of eclipses from their data. Now we live in a society of specialists.)

The symbiosis spins as a cycle; the experimentalist discovers new data; the theorist must then

---

Supported in part by Harvard Society of Fellows and the National Science Foundation under Grant PHY79-16812.

deduce a new pattern that fits the data. The theorist predicts from the pattern phenomena for which the experimentalist probes, the probes producing again new data.

The theoretical physicist describes the patterns he sees with mathematics. In order to predict new phenomena, a mathematical model is made. Here, one must think, in the simplest example, of a simple polynomial expression: $y = ax + b$. This expression expresses the output, called y, in terms of the input, called x, and parameters, called a and b. The experimentalist inputs a string of x's and measures with machines the values of y. Later, the theorist comes to conclude that y is related to x, as given by the equation above, with "a" and "b" as "such and such," and that when the experimentalist tries $x =$ "*doo*," he will surely record that $y =$ "*daa*."

If the theoretician truly suspects that there is a polynomial modeling our universe, then he will certainly learn as much as he can about polynomial equations. The knowledge allows him to publish predictions that will test the power of the proposed polynomial model.

Before you realize it, you have a physicist who is doing abstract algebraic geometry. Indeed, one may argue that such a scenario certainly suggests how the human race started to study geometry in the first place. The point I am pushing is that theoretical physics and modern mathematics are mutually entwined. Indeed, it has seemed, at least in the past, that finding new physics leads to learning (or inventing) new math.

If I may return to my recent example: when the measuring machine, matching x's with y's starts to suggest a sinusoidal situation, then the theoretical physicist had better discover trigonometric functions, or else retire.

Mathematics is a tool for the physicist and for this reason a student of physics should be encouraged to learn the elegant formulations of his physical theories. (There is always the danger of some student finding the tool more tantalizing than the physics, but the world needs young mathematicians as well.) Professor Felsäger has written a book that will introduce seniors and new graduate students in physics to the elegant mathematics one discovers while studying electro-magnetic phenomena and quantum mechanics.

The study of electro-magnetism is the study of electric fields, magnetic fields, and the forces they put on passing particles. The subject finds phenomena from fluorescent lights to lightning bolts to aurora borealis and electric eels.

The mathematical description of the electro-magnetic field is most elegantly elucidated by the language of modern differential geometry and topology. This is the language that Professor Felsäger teaches. But he presents more than a language; for a new language opens a new window with which to view the physical world.

The language of differential geometry allows one to ask new questions which were not even poseable before. There is such a question whose answer may come to be the fundamental discovery of this half century. The question and the reasoning that led to its asking is readily recounted. Recall from childhood playing with magnets. They came in all sizes, colors: bar magnets, and horseshoe magnets. But no matter which, they all had two ends; all of mine were marked on one end with an N and the other with an S: the north pole and the south pole. Poles of the same kind repel each other while unlike poles attract. I never saw a magnet with only one pole. In fact, every magnet ever encountered or built has two poles; one north and the other south. The mathematics of electromagnetism is the mathematics of line bundles, covariant derivatives, and curvature. This dipole phenomenon is described completely by the differential geometry of line bundles; but remarkably the phenomenon is specific to line bundles, as opposed to vector bundles of higher rank.

Thinking about differential geometry led physicists to speculate about the nature of a world whose forces were modeled on vector bundles of higher rank. Indeed, such vector bundle models would better describe our world. There are forces in nature other than electricity and magnetism and these models unify all the forces of nature as facets of one object: a *connection* on a vector bundle of high rank. This is a beautiful proposition. But if this Grand Unified Theory, as it is called, is an accurate model, there most certainly should exist magnetic monopoles. These are the

as yet unencountered pieces of matter with just one pole, either north or south. The critical question is whether there is or is not a real monopole out there, somewhere. The experimental physicists are strenuously seeking it, or its telltale traces.

The second subject of Professor Felsäger's book is quantum mechanics. This is treated with considerably less rigor, and perhaps reasonably so: it is a much more difficult subject. Indeed, the welding of quantum mechanics with electro-magnetism (called QED, quantum-electrodynamics) by a well-defined mathematical model has eluded the theorists today. Theoretical physicists have assembled an algorithm to calculate quantities in QED. The quantities calculated using the algorithm are amazingly accurate. But today there is little basis for believing the algorithm except that it has never yet been wrong. Is there a mathematical model behind it? Investigations have suggested an interesting hypothesis: only theories based on the differential geometry of vector bundles of high rank can be welded with quantum mechanics in a rigorous mathematical model. The suggestion is that Grand Unified Theories are accurate because there is no way to build any other kind.

The construction of the quantum Grand Unified Theories is a formidable task which is only just begun. The algorithms for calculating physical parameters are extremely complicated. When one considers the Grand Unified Theories, the nonlinearities in the theory are of primary importance, and so far, rigorous results are rare.

These mathematical models are so monstrous that physicists do not know what they predict. The physicist has only his intuition as a guide, a large part of which comes through studying the differential geometric aspects of these theories.

The application of differential geometry to calculate quantum amplitudes is mathematically justified in simple models; and Professor Felsäger presents a number of these calculations. From a practical viewpoint, these calculations are the most useful parts of his book, because the intuition for Grand Unified Theories comes from studying the role of differential geometry in the simple situations. Here I have my major criticism of Professor Felsäger's book: he does not give enough references. The format is expository without providing the rigorous background to quantum field theory. I think the author should have given us a large selection of references to constructive quantum field theory and to the many "instanton" papers and calculations using differential geometry. It defeats the purpose to present this thought-provoking subject without also telling the reader where to find more detail. Except for this defect, Professor Felsäger's book is wholly commendable.

An Introduction to Classical Real Analysis. By Karl R. Stromberg. Wadsworth, Belmont, California, 1981. ix + 576 pp., $29.85.

ALBERTO TORCHINSKY
Department of Mathematics, Indiana University, Bloomington, IN 47405

Skip the bills, exams, reviews and other distractions at hand, play one of Hindemith's wood-wind sonatas (the Beatles will also do), and, most importantly, read *An Introduction to Classical Real Analysis*, Karl Stromberg's expert work of art and love, in the basic subject described by its title. Stromberg condenses over twenty years of teaching experience in an attractive treatise meant for those students who, having survived our merciful attempts to teach them Calculus without ε's and δ's, want to know how things can be really made to work. Such a book is especially welcome at this introductory level, for the main texts in print are the second editions of Apostol and Rudin, written in the midseventies. Both of these authors made a significant contribution to the study of Classical Analysis, or Advanced Calculus as the subject is sometimes referred to, and their highly successful texts do provide a solid preparation for

as yet unencountered pieces of matter with just one pole, either north or south. The critical question is whether there is or is not a real monopole out there, somewhere. The experimental physicists are strenuously seeking it, or its telltale traces.

The second subject of Professor Felsäger's book is quantum mechanics. This is treated with considerably less rigor, and perhaps reasonably so: it is a much more difficult subject. Indeed, the welding of quantum mechanics with electro-magnetism (called QED, quantum-electrodynamics) by a well-defined mathematical model has eluded the theorists today. Theoretical physicists have assembled an algorithm to calculate quantities in QED. The quantities calculated using the algorithm are amazingly accurate. But today there is little basis for believing the algorithm except that it has never yet been wrong. Is there a mathematical model behind it? Investigations have suggested an interesting hypothesis: only theories based on the differential geometry of vector bundles of high rank can be welded with quantum mechanics in a rigorous mathematical model. The suggestion is that Grand Unified Theories are accurate because there is no way to build any other kind.

The construction of the quantum Grand Unified Theories is a formidable task which is only just begun. The algorithms for calculating physical parameters are extremely complicated. When one considers the Grand Unified Theories, the nonlinearities in the theory are of primary importance, and so far, rigorous results are rare.

These mathematical models are so monstrous that physicists do not know what they predict. The physicist has only his intuition as a guide, a large part of which comes through studying the differential geometric aspects of these theories.

The application of differential geometry to calculate quantum amplitudes is mathematically justified in simple models; and Professor Felsäger presents a number of these calculations. From a practical viewpoint, these calculations are the most useful parts of his book, because the intuition for Grand Unified Theories comes from studying the role of differential geometry in the simple situations. Here I have my major criticism of Professor Felsäger's book: he does not give enough references. The format is expository without providing the rigorous background to quantum field theory. I think the author should have given us a large selection of references to constructive quantum field theory and to the many "instanton" papers and calculations using differential geometry. It defeats the purpose to present this thought-provoking subject without also telling the reader where to find more detail. Except for this defect, Professor Felsäger's book is wholly commendable.

An Introduction to Classical Real Analysis. By Karl R. Stromberg. Wadsworth, Belmont, California, 1981. ix + 576 pp., $29.85.

ALBERTO TORCHINSKY
Department of Mathematics, Indiana University, Bloomington, IN 47405

Skip the bills, exams, reviews and other distractions at hand, play one of Hindemith's wood-wind sonatas (the Beatles will also do), and, most importantly, read *An Introduction to Classical Real Analysis*, Karl Stromberg's expert work of art and love, in the basic subject described by its title. Stromberg condenses over twenty years of teaching experience in an attractive treatise meant for those students who, having survived our merciful attempts to teach them Calculus without ε's and δ's, want to know how things can be really made to work. Such a book is especially welcome at this introductory level, for the main texts in print are the second editions of Apostol and Rudin, written in the midseventies. Both of these authors made a significant contribution to the study of Classical Analysis, or Advanced Calculus as the subject is sometimes referred to, and their highly successful texts do provide a solid preparation for

undergraduate and beginning graduate students learning complex and abstract analysis. In fact, the clarity of their presentation affords in a sense an invitation to the students to "read through" the book. Stromberg's book, on the other hand, while just as clear as those of Apostol and Rudin, advances a somewhat different point of view: the author insists that to learn Mathematics (and this subject in particular) hard work is necessary, and he incorporates into the flow of the presentation the solving of sets of exercises, some supplied with copious hints. Most of the exercises represent assertions to be proved and lead the students to important and interesting results; this is an attractive asset.

Readers will have no difficulty in deciding that in the balance of rigor vs. intuition, the author has chosen an extremely rigorous approach which opens with a set of axioms introducing the real numbers and culminates with a discussion of the growth rate of partial sums of Fourier series. A safe assumption is made, namely that students know no mathematics, and a point of view is advanced, namely that no concept will be used until it is properly introduced; thus π is not mentioned until Chapter 5. The emphasis (and in some parts, as in the convergence tests for numerical series, it is heavily laid) is on the *classical* aspect of the theory. For instance, although the concepts of abstract topological and metric spaces are indeed introduced in Chapter 1, the book promptly reverts to real Euclidean space for examples, motivation, and development. Complex valued functions are also considered, but the theory of analytic functions is not discussed. As for functions of several real variables, although the Fubini-Tonelli theorem and formulas such as the change of variables involving Jacobians are given, the theorems of Stokes and Green are not covered. Students in engineering and other applied areas will be pleasantly surprised by a significant simplification, if that is the right word, which Stromberg proposes. The Lebesgue integral is introduced before other notions of integral. To support this concept Stromberg advances three reasons; I add two of my own: it saves time and it works well.

The different chapters that comprise the book can be read independently, the exercises are challenging and fun, and the content is supported by the clear, concrete, and correct mathematical thinking (in both level and presentation) that is such a crucial component of the book. Through the chapters students can learn about the number system, sequences, and series, limits and continuity, differentiation, the elementary transcendental functions, integration (and its main applications including L_p spaces, and some differential calculus in several real variables), infinite series and products, and trigonometric series. Stromberg indicates to the prospective teacher how to program a two-semester, or three-quarter, course by marking in the index with a single asterisk sections that may be omitted if time presents a problem and with a double asterisk interesting and useful applications of the theory.

I feel this book will become a source of inspiration for teachers and students of classical analysis. It has become one of my reference books and one of my first choices for the kind of course where students want to see the proofs of the theorems, and consequently the whole theory, come alive.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

This section seems an appropriate place to comment on a recent "conjecture" by the editor. I refer to the "conjecture" described by the title of Professor Halmos's recent essay, "Applied Mathematics Is Bad Mathematics," in *Mathematics Tomorrow* [**1**]. In the event you've missed it, in the lead article of a section titled "What is Mathematics?" Professor Halmos has decided to

undergraduate and beginning graduate students learning complex and abstract analysis. In fact, the clarity of their presentation affords in a sense an invitation to the students to "read through" the book. Stromberg's book, on the other hand, while just as clear as those of Apostol and Rudin, advances a somewhat different point of view: the author insists that to learn Mathematics (and this subject in particular) hard work is necessary, and he incorporates into the flow of the presentation the solving of sets of exercises, some supplied with copious hints. Most of the exercises represent assertions to be proved and lead the students to important and interesting results; this is an attractive asset.

Readers will have no difficulty in deciding that in the balance of rigor vs. intuition, the author has chosen an extremely rigorous approach which opens with a set of axioms introducing the real numbers and culminates with a discussion of the growth rate of partial sums of Fourier series. A safe assumption is made, namely that students know no mathematics, and a point of view is advanced, namely that no concept will be used until it is properly introduced; thus π is not mentioned until Chapter 5. The emphasis (and in some parts, as in the convergence tests for numerical series, it is heavily laid) is on the *classical* aspect of the theory. For instance, although the concepts of abstract topological and metric spaces are indeed introduced in Chapter 1, the book promptly reverts to real Euclidean space for examples, motivation, and development. Complex valued functions are also considered, but the theory of analytic functions is not discussed. As for functions of several real variables, although the Fubini-Tonelli theorem and formulas such as the change of variables involving Jacobians are given, the theorems of Stokes and Green are not covered. Students in engineering and other applied areas will be pleasantly surprised by a significant simplification, if that is the right word, which Stromberg proposes. The Lebesgue integral is introduced before other notions of integral. To support this concept Stromberg advances three reasons; I add two of my own: it saves time and it works well.

The different chapters that comprise the book can be read independently, the exercises are challenging and fun, and the content is supported by the clear, concrete, and correct mathematical thinking (in both level and presentation) that is such a crucial component of the book. Through the chapters students can learn about the number system, sequences, and series, limits and continuity, differentiation, the elementary transcendental functions, integration (and its main applications including L_p spaces, and some differential calculus in several real variables), infinite series and products, and trigonometric series. Stromberg indicates to the prospective teacher how to program a two-semester, or three-quarter, course by marking in the index with a single asterisk sections that may be omitted if time presents a problem and with a double asterisk interesting and useful applications of the theory.

I feel this book will become a source of inspiration for teachers and students of classical analysis. It has become one of my reference books and one of my first choices for the kind of course where students want to see the proofs of the theorems, and consequently the whole theory, come alive.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

This section seems an appropriate place to comment on a recent "conjecture" by the editor. I refer to the "conjecture" described by the title of Professor Halmos's recent essay, "Applied Mathematics Is Bad Mathematics," in *Mathematics Tomorrow* [**1**]. In the event you've missed it, in the lead article of a section titled "What is Mathematics?" Professor Halmos has decided to

address himself to this question by explaining that we have Pure Mathematics (PM) and Applied Mathematics (AM), and that the latter is "usually" "bad mathematics" (BM), i.e., AM = BM a.e. PM is not explicitly labeled "good," but the reader is unlikely to be confused by the omission.

Actually, we never do find out very clearly just what "bad" means (vague talk of standards of clarity, elegance, precision), although a sufficient number of self-serving examples of Pure vs. Applied categories is supplied to allow us to make some inferences; these are on the level of: Knowing vs. Doing, Mozart vs. military marches, Beauty vs. Boredom, and so on. Guess which is which. What we do get seems consciously coy (a kind of mathematical I'm-just-telling-it-like-it-is-folks), and largely an exhumation of a point of view amply spoken for by Hardy in *A Mathematician's Apology* [**2**]. (Not everyone took Hardy's "apology" seriously, and the review of [**2**] by F. R. Soddy [**3**] is well worth looking up; he refers to [**2**] as "such cloistral clowning," of which "AM is BM" also contains its share.)

What's the point? After all, von Neumann (no slouch in either AM or PM) in commenting on these matters in "The Mathematician" (in Newman [**4**]) has long ago resolved the issue in a realistic, useful, and positive way, has he not? Von Neumann stresses the "given" character of applied problems, their concentrated and circumscribed nature, and the fact that an important AM problem represents a difficulty which "must" be resolved—there is often no opportunity to exercise aesthetic criteria of selection and success. He also seems uninterested in promulgating mathematicians' myths of their own power and taste, and I dare say he might give the answer which can be given to much of the kind of "good-bad" discussion in "AM is BM": Try it.

The intermittent sallies by those in PM who would like to dissociate themselves from AM (or vice versa), while not always enlightening, often make livelier than usual professional reading. I imagine "AM is BM" was fun to write, and, for one opinion of "What is Mathematics?" it's worth reading. How about the MONTHLY running a series of articles by others addressing the same question, the editor to work out the details?

1. L. A. Steen, editor, Mathematics Tomorrow, Springer, NY, 1981.
2. G. H. Hardy, A Mathematician's Apology, Cambridge University Press, 1977.
3. F. R. Soddy, Qui s'accuse s'acquitte, Nature, Jan. 4, 1941, 3–5.
4. J. R. Newman, editor, The World of Mathematics, vol. 4, Simon and Schuster, NY, 1956.

Richard C. Bollinger
Behrend College, Pennsylvania State University
Erie, PA 16563

MISCELLANEA

Continuity: A Do-it-yourself Editorial

Suggested by S. I. Gass and R. C. Buck; titled and prepared for publication by R. P. Boas.

102. Here are some definitions of continuity, arranged chronologically. Draw your own conclusions.

1. (1817) According to a correct definition, we are to understand only this much by saying that

address himself to this question by explaining that we have Pure Mathematics (PM) and Applied Mathematics (AM), and that the latter is "usually" "bad mathematics" (BM), i.e., AM = BM a.e. PM is not explicitly labeled "good," but the reader is unlikely to be confused by the omission.

Actually, we never do find out very clearly just what "bad" means (vague talk of standards of clarity, elegance, precision), although a sufficient number of self-serving examples of Pure vs. Applied categories is supplied to allow us to make some inferences; these are on the level of: Knowing vs. Doing, Mozart vs. military marches, Beauty vs. Boredom, and so on. Guess which is which. What we do get seems consciously coy (a kind of mathematical I'm-just-telling-it-like-it-is-folks), and largely an exhumation of a point of view amply spoken for by Hardy in *A Mathematician's Apology* [**2**]. (Not everyone took Hardy's "apology" seriously, and the review of [**2**] by F. R. Soddy [**3**] is well worth looking up; he refers to [**2**] as "such cloistral clowning," of which "AM is BM" also contains its share.)

What's the point? After all, von Neumann (no slouch in either AM or PM) in commenting on these matters in "The Mathematician" (in Newman [**4**]) has long ago resolved the issue in a realistic, useful, and positive way, has he not? Von Neumann stresses the "given" character of applied problems, their concentrated and circumscribed nature, and the fact that an important AM problem represents a difficulty which "must" be resolved—there is often no opportunity to exercise aesthetic criteria of selection and success. He also seems uninterested in promulgating mathematicians' myths of their own power and taste, and I dare say he might give the answer which can be given to much of the kind of "good-bad" discussion in "AM is BM": Try it.

The intermittent sallies by those in PM who would like to dissociate themselves from AM (or vice versa), while not always enlightening, often make livelier than usual professional reading. I imagine "AM is BM" was fun to write, and, for one opinion of "What is Mathematics?" it's worth reading. How about the MONTHLY running a series of articles by others addressing the same question, the editor to work out the details?

1. L. A. Steen, editor, Mathematics Tomorrow, Springer, NY, 1981.
2. G. H. Hardy, A Mathematician's Apology, Cambridge University Press, 1977.
3. F. R. Soddy, Qui s'accuse s'acquitte, Nature, Jan. 4, 1941, 3–5.
4. J. R. Newman, editor, The World of Mathematics, vol. 4, Simon and Schuster, NY, 1956.

Richard C. Bollinger
Behrend College, Pennsylvania State University
Erie, PA 16563

MISCELLANEA

Continuity: A Do-it-yourself Editorial

Suggested by S. I. Gass and R. C. Buck; titled and prepared for publication by R. P. Boas.

102. Here are some definitions of continuity, arranged chronologically. Draw your own conclusions.

1. (1817) According to a correct definition, we are to understand only this much by saying that

a [real-valued] function $f(x)$ varies according to the law of continuity, for all values of x between or outside certain bounds: that if x is any such value, the difference $f(x+\omega) - f(x)$ can be made smaller [in absolute value] than any given number, provided that we can take ω as small as we please.

—Bolzano [**1**], pp. 7–8

[That the words in brackets were intended is clear from the context and from other writings of Bolzano; cf. [**7**]. Bolzano, like many generations of physical scientists and college freshmen, uses "small" to mean "close to zero."]

2. (1857) A continuous function is subject to the two following assumptions: 1st. As the variable gradually changes, the function must gradually change. 2nd. The law symbolized by the functional character must not abruptly change. When these two conditions are not satisfied, the function is discontinuous.

—Price [**6**], p. 28

3. (1863) A function is continuous, when it undergoes a gradual change; it is discontinuous, when the change is not gradual, or when the function changes suddenly from one value to another very different value. Thus when the difference between $f(x)$ and $f(x+h)$ may, by the continued diminuition of h, be made as small as we please, $f(x)$ is a continuous function; but when under the same circumstances $f(x+h)$ differs widely from $f(x)$, the latter is a discontinuous function.

—Hall [**3**], p. 2

4. (1891) A function is said to be continuous between certain values of the independent variable, when it changes *gradually* while the variable passes from one value to the other. In other words, a *continuous* function is one that can be represented by a *continuous* curve.

—Osborne [**4**], p. 64

5. (1922) A function, $f(x)$, is said to be continuous if a slight change in x produces but a slight change in the value of the function.... We can now make more explicit the definition given in Chapter I by saying: $f(x)$ is continuous at the point $x = a$ if $\lim_{x \doteq a} f(x) = f(a)$.

—Osgood [**5**], pp. 7, 28

6. (1978) A numerical-valued function f, defined on a set D, is said to be continuous at a point $p_0 \in D$ if, given any number $\varepsilon > 0$, there is a neighborhood U about p_0 such that $|f(p) - f(p_0)| < \varepsilon$ for every point $p \in U \cap D$. The function f is said to be continuous on D if it is continuous at each point of D.

—Buck [**2**], p. 72

References

1. Bernard Bolzano, Rein analytischer Beweis des Lehrsatzes, dass zwischen je zwey Werther, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege, Prague, 1817; reproduced in Ostwald's Klassiker der exakten Wissenschaften, no. 153, Leipzig, 1905.

2. R. C. Buck and E. F. Buck, Advanced Calculus, 3rd ed., McGraw-Hill, New York, 1978.

3. T. G. Hall, A Treatise on the Differential and Integral Calculus, 6th ed., Longman, London, 1863.

4. G. A. Osborne, An Elementary Treatise on the Differential and Integral Calculus, Leach, Shewell, and Sanborn, Boston-New York-Chicago, 1891.

5. W. F. Osgood, Introduction to the Calculus, Macmillan, New York, 1922.

6. Bartholomew Price, A Treatise on Infinitesimal Calculus, 2nd ed., Oxford, 1857.

7. O. Stolz, B. Bolzano's Bedeutung in der Geschichte der Infinitesimalrechnung, Math. Ann., 18 (1881) 255–279.

new math texts from

Manifolds, Tensor Analysis, and Applications

Global Analysis Series, No. 2*

Ralph Abraham, *University of California, Santa Cruz,* **Jerrold E. Marsden,** *University of California, Berkeley,* and **Tudor Ratiu,** *University of Michigan*

This text, providing core background material in global analysis and a working knowledge of manifolds, dynamical systems, tensors, and differential forms, is suitable as a textbook for the graduate student as well as a reference book for the established researcher.

1983, approx. 575 pp., illus. with line drawings, halftones
Hardbound ISBN 0-201-10168-8 $34.95

Real Analysis

Second Edition, revised, enlarged, reset

Serge Lang, *Yale University*

Designed for a basic graduate course in real analysis, this textbook covers the fundamentals of measure and integration theory, and of functional analysis. Lang has incorporated the suggestions of users of the first edition to make this second edition an even more useful textbook for beginning graduate students, containing more exercises than the first, and including concrete applications of the general theory.

1983, 533 pp., illus. with line drawings
Hardbound ISBN 0-201-14179-5 $23.95

Elements of Algebra and Algebraic Computing

John D. Lipson, *University of Toronto*

By combining the traditional material of modern and universal algebra with the relatively new ideas of algebraic computing, this book provides a view of algebra influenced by the computer which is suitable for one- or two-semester applied algebra courses.

1981, 356 pp., illus. with line drawings
Hardbound ISBN 0-201-04115-4 $34.50

Algebraic Topology: A First Course

Marvin J. Greenberg, *University of California, Santa Cruz,* and **John R. Harper,** *University of Rochester*
Mathematics Lecture Notes Series, Vol. 58*

This revision of Marvin J. Greenberg's Lectures on Algebraic Topology has added theory, examples, and exercises to update, enhance, and simplify the original exposition, while retaining the point of view and organizational principles of the earlier book.

1981, 250 pp., illus. with line drawings
Hardbound ISBN 0-8053-3558-7 $36.95
Paperbound ISBN 0-8053-3557-9 $19.50

forthcoming . . . Harmonic Analysis

Henry Helson, *University of California, Berkeley*

from the author's preface: "The basic examples and classical results of all parts of analysis, and beyond, are to be found in the beautiful subject that used to be called Fourier series, and now goes by the more exalted name of Harmonic Analysis. Even if the classical techniques are not sufficient to solve modern problems, a knowledge of them is indispensable for a proper view of the methods that have descended from them . . . To be sure there is a big gap between learning the Banach-Steinhaus theorem, for example, and applying it to a real problem. Filling that gap is one of the objectives of this book."

1983, approx. 175 pp.
Hardbound ISBN 0-201-12752-0 tentative price $22.95

Prices are quoted in U S. dollars and are subject to change without notice. Prices vary outside the U.S.A
*continuation orders invited

ADVANCED BOOK PROGRAM

Addison-Wesley Publishing Company, Inc.
Advanced Book Program/World Science Division
Reading, Massachusetts 01867, U.S.A.

6 Byfield Street, North Ryde, N.S.W. 2113, Australia • 53 Bedford Square, London WCIB 3DZ, England • De Lairessestraat 90, 1071 PJ Amsterdam, Netherlands • 36 Prince Andrew Place, Don Mills, Ontario M3C 2T8, Canada • I.P.O. Box 5015, Tokyo 100-31, Japan

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 90, Number 5 May 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

TRICKS OR TREATS WITH THE HILBERT MATRIX

MAN-DUEN CHOI

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 1A1

This article is intended to be a partial account on everything you always wanted to know about the *Hilbert matrix*—namely, the infinite Hilbert matrix A or the finite order $n \times n$ Hilbert matrix A_n:

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & & \\ \frac{1}{3} & \frac{1}{4} & & & \\ \frac{1}{4} & & & & \\ \vdots & & & & \end{bmatrix}, \quad A_n = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & & & \\ \frac{1}{3} & & & & \\ \vdots & & & & \vdots \\ \frac{1}{n} & & & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

It is natural to ask: What are the natural consequences from such a natural assemblage of (the reciprocals of) natural numbers? Hence, here arise ten concrete *problems* (instead of theorems), aimed to reveal the "heart of mathematics" in a broad sense. All solutions provided in this article are elementary (requiring only the rudiments of linear algebra or simple operator theory without the spectral theorem). However, many tricks or treats associated with the Hilbert matrix may seem rather frightening or fascinating.

The Hilbert matrix has played a prominent role in the structure theory of several branches of mathematics. Indeed, it serves as one of the most vivid examples for many unusual aspects (as well as usual aspects, of course) in operator theory (see e.g., **[1]**, **[4]**, **[7]**). Undoubtedly, the infinite Hilbert matrix stands out from other single operators, illustrating many soft results by hard analysis, discrete features of continuous phenomena, and subtle effects through intricate combinatorics.

This pseudo-expository article is designed to exhibit some of these phenomena. The main body consists of ten problems on the Hilbert matrix, showing various aspects of concrete and discrete nature. The majority of these problems are apparently new, but many of their variants and underpinnings have already appeared in the literature. The second half of the article is devoted to solutions and additional notes and references. Much effort has been made to seek the most elementary solutions instead of the best answers to the problems. Interested readers will probably want to find other solutions that are less elementary but provide still more information.

It may be appropriate to fix some notations and terms here. The term "matrix" refers to a finite-order or infinite square matrix with complex entries (or real entries, without loss of generality). The letter I stands for the identity matrix with ones along the main diagonal and zeros elsewhere. A matrix $T = [\tau_{ij}]_{i,j}$ is said to be *positive*, in notation $T \geqslant 0$, if $\sum \tau_{ij}\bar{x}_i x_j \geqslant 0$ for all complex k-tuples $(x_1, x_2, \ldots, x_k)$ with $k \geqslant 0$. The notation $\|T\|$ refers to the *Hilbert-space-operator norm*; thus if an operator T: $l^2 \to l^2$ has the matrix representation $[\tau_{ij}]$, then the operator norm of T (that is, the operator norm of $[\tau_{ij}]$) is

Man-Duen Choi was brought up in Hong Kong. He received his Ph.D. degree under the guidance of Chandler Davis at Toronto. He taught at the University of California, Berkeley, from 1973 to 1976, and has worked since then at the University of Toronto. His research has been mainly in operator algebras, operator theory, and polynomial rings. He is particularly interested in examples/counterexamples and two-by-two matrix manipulations.

$$\|T\| = \sup\left\{\left(\sum_i\left|\sum_j \tau_{ij}\alpha_j\right|^2\right)^{1/2} : \sum|\alpha_j|^2 \leqslant 1\right\}.$$

There is a useful elementary fact: if $T = [\tau_{ij}]$ and $S = [\sigma_{ij}]$ with $0 \leqslant \sigma_{ij} \leqslant \tau_{ij}$ for all i and j, then $\|S\| \leqslant \|T\|$.

I. Invertibility. By computation, $A_1^{-1} = 1$,

$$A_2^{-1} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix}, \quad A_3^{-1} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

Thus, the general case becomes irresistible.

PROBLEM: *Write out the inverse of the $n \times n$ Hilbert matrix A_n explicitly.*

It may be instructive to expose what you were afraid to ask:

(a) While each entry of A_n is the reciprocal of an integer, what sort of coincidence is it if A_n^{-1} has integer entries?

(b) In particular, why is det A_n the reciprocal of an integer?

(c) Since each entry of A_n is a positive real number, does it follow that A_n^{-1} is of the form $[\beta_{ij}]$ with $(-1)^{i+j}\beta_{ij} \geqslant 0$ for all i, j?

(d) While A_n is a *Hankel matrix* (i.e., $A_n = [\alpha_{ij}]$ with $\alpha_{ij} = \alpha_{pq}$ whenever $i + j = p + q$), should there be any pattern for A_n^{-1} at all?

II. Formal Inverse. With all the tractable A_n^{-1} already in hand, can A^{-1} be still far beyond the grasp?

In this section, the notation Σ is used only for sequential summation—absolute convergence is not required. There are three intrinsic properties pertinent to each infinite matrix $T = [\tau_{ij}]$: T is said to be *formally one-to-one* if the trivial sequence is the only sequence (α_j) satisfying $\Sigma_j \tau_{ij}\alpha_j = 0$ for each i; T is said to be *formally onto* if for each sequence (α_j) there is a sequence (β_j) such that $\Sigma_j \tau_{ij}\alpha_j = \beta_i$ for each i; $S = [\sigma_{ij}]$ is called a *formal inverse* of T if the formal products TS and ST are defined and equal to I; that is,

$$\sum_k \tau_{ik}\sigma_{kj} = \sum_k \sigma_{ik}\tau_{kj} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Since the formal multiplication of infinite matrices is nonassociative, there is no actual correlation among these three properties.

PROBLEM: (1) *Is the infinite Hilbert matrix A formally one-to-one?*

(2) *Is A formally onto?*

(3) *Does A have a formal inverse?*

III. Strong Positivity. If $T = [\tau_{ij}]$ is a matrix with $\tau_{ij} \geqslant 0$ and r is a real number, then let $T^{[r]}$ denote the matrix $[\tau_{ij}^r]$.

It is worthwhile to praise the $n \times n$ Hilbert matrix A_n for all its positivity in the highest.

PROBLEM: *Prove that $A_n^{[r]} \geqslant 0$ for each real number $r \in (0, \infty)$.*

Note that in general, "$T \geqslant 0$" need not imply "$T^{[r]} \geqslant 0$"; example:

$$T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \geqslant 0, \quad \text{but } T^{[1/2]} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2^{1/2} & 1 \\ 0 & 1 & 1 \end{bmatrix} \ngeqslant 0.$$

IV. Factorization. In view of the positivity of the infinite Hilbert matrix, Paul Halmos (oral communication) has asked: Is it possible to write out explicitly $A = B^2$ with $B \geqslant 0$? This question looks very frightening—apparently, it may involve finding the spectral decomposition of A. Still,

there is a more tractable problem.

PROBLEM: *Write out explicitly* $A = BB^*$ *where* B *is lower-triangular* (i.e., $B = [\beta_{ij}]_{1 \leqslant i, j < \infty}$ with $\beta_{ij} = 0$ whenever $i < j$). (Note that such B is unique if, in addition, all diagonal entries of B are positive real numbers.)

There are some merits for the triangular factorization:

(a) Let $B_n = [\beta_{ij}]_{1 \leqslant i, j \leqslant n}$ be the $n \times n$ upper-left corner of B. Then it follows $A_n = B_n B_n^*$. Thus an induction to seek B is admissible.

(b) Furthermore, $\det A_n = |\det B_n|^2$ is completely determined by the diagonal entries of B. (Cf. remark (b) following the problem in Section I.)

(c) Because of its very tractable structure, each lower-triangular infinite matrix with nonzero diagonal entries always admits a formal inverse which is also lower-triangular. Thus let C be the formal inverse of B. Then the abstract expression $C^*C = A^{-1}$ makes sense in the following manner: If $Av = w$, then $C^*(Cw) = v$—the formal matrix manipulations for Cw and $C^*(Cw)$ can be carried out even though the formal matrix product C^*C is not defined.

V. Companion. A loyal companion of the Hilbert matrix is the matrix

$$L = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \\ \vdots & & & \end{bmatrix} \quad \text{or } L_n = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{2} & & \\ \vdots & & & \vdots \\ \frac{1}{n} & & \cdots & \frac{1}{n} \end{bmatrix}$$

whose entries run in a reversed-L-shaped pattern.

Notice that L (or L_n) is an assemblage of reciprocals of positive integers, too. Aside fr(external resemblance, L (or L_n) also enjoys all those delightful properties as A (or A_n) does in the preceding problems. Notably, L admits a triangular factorization $L = CC^*$ where C is the infinite Cesàro matrix,

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \\ \vdots & & & & \end{bmatrix}.$$

It is very pleasant to call upon the companion L for the proof that $A : l^2 \to l^2$ is a bounded operator.

PROBLÉM: (1) *Observe that* $I - (I - C)(I - C)^* \geqslant 0$.

(2) *Prove that* $\|A\| \leqslant \|L\| \leqslant 4$.

Apparently, this problem provides the quickest elementary way to show that A is a bounded operator.

VI. Heredity. The immediate offspring of the infinite Hilbert matrix A is

$$A' = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots \\ \frac{1}{3} & \frac{1}{4} & & \\ \frac{1}{4} & & & \\ \vdots & & & \end{bmatrix}$$

which has shared the name of *Hilbert matrix* in literature.

Undoubtedly, A' deserves the title of Hilbert matrix because A' indeed carries on all heredity.

PROBLEM: *Show that A and A' are unitarily equivalent* (i.e., there exists a unitary operator U such that $A = U^*A'U$).

It is apparently surprising to see that (a) A' is entrywise strictly smaller than A, and (b) A' is a proper part of A—the first row (or the first column) of A has been erased, yet it turns out that A' is unitarily equivalent to A.

VII. Compact Perturbation. Many offspring of the infinite Hilbert matrix A can be obtained by simply erasing the first p_1 rows and the first p_2 columns from A. Thus, there are infinitely many Hankel matrices

$$A(q) = \begin{bmatrix} \frac{1}{q+1} & \frac{1}{q+2} & \frac{1}{q+3} & \cdots \\ \frac{1}{q+2} & \frac{1}{q+3} & & \\ \frac{1}{q+3} & & & \\ \vdots & & & \end{bmatrix}$$

with $q = p_1 + p_2$ being any nonnegative integer. In particular, $A(0) = A$, $A(1) = A'$.

It is a pleasure to report below that all these offspring are *essentially the same* (i.e., the difference of any two of them is a compact operator). Let $T_n = [\tau_{ij}]_{1 \leqslant i, j \leqslant n}$ be the $n \times n$ upper-left corner matrix of $T = [\tau_{ij}]_{1 \leqslant i, j < \infty}$; T_n can be regarded as an infinite matrix $[\alpha_{ij}]_{1 \leqslant i, j < \infty}$ with

$$\alpha_{ij} = \begin{cases} \tau_{ij} & \text{if } i \leqslant n \text{ and } j \leqslant n, \\ 0 & \text{elsewhere.} \end{cases}$$

By one of many equivalent definitions, T is said to be *compact* if $\lim_{n\to\infty}\|T - T_n\| = 0$; T is said to be of *trace-class* if the sum of the main diagonal entries of $(T^*T)^{1/2}$ is finite. It is well known that

$$\{\text{trace-class operators}\} \subset \{\text{compact operators}\} \subset \{\text{bounded operators}\}$$

where $\subset$ stands for a proper inclusion as a linear submanifold.

PROBLEM: (1) *Show that the infinite Hilbert matrix A is not a compact operator.*

(2) *Show that $A - A(q)$ is a trace-class operator.*

VIII. The Euler Recipe. There are many possible ways to make π; but it is striking to get π from the magnificent formula of Euler

$$\pi^2/6 = 1 + 1/2^2 + 1/3^2 + 1/4^2 + \cdots .$$

A countable ordered set Γ is called an *Euler vector* if Γ as an unordered set is the same as $\{0\} \cup \{1/k\colon k \in \mathbb{Z} \setminus \{0\}\}$, and furthermore each nonzero element $1/k$ occurs three times exactly in Γ. By Euler's formula, it follows immediately that each Euler vector is of norm π in the l^2-norm.

Next, recall that a matrix T_1 is a *dilation* of a matrix T_0 if T_1 can be written in the form $\begin{bmatrix} X & Y \\ Z & T_0 \end{bmatrix}$. In other words, if $\mathcal{H}_0$ is a subspace of a Hilbert space $\mathcal{H}_1$ and P is the orthogonal projection from $\mathcal{H}_1$ onto $\mathcal{H}_0$, and if $T_1\colon \mathcal{H}_1 \to \mathcal{H}_1$, $T_0\colon \mathcal{H}_0 \to \mathcal{H}_0$ are operators satisfying $PT_1|_{\mathcal{H}_0} = T_0$, then T_1 is called a *dilation* of T_0.

PROBLEM: *Dilate the infinite Hilbert matrix A (or its immediate offspring A') to a real symmetric matrix T such that each column-vector of T is an Euler vector, and furthermore any two column-vec-*

tors of T are orthogonal to each other.

Consequently, $T^2 = \pi^2 I$ and $\|T\| = \pi$; therefore $\|A\| \leqslant \|T\| = \pi$.

IX. π Again. A close relative of the $n \times n$ Hilbert matrix A_n is

$$Z_n = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & & & 0 \\ \frac{1}{3} & & & & \\ \vdots & & & & \vdots \\ \frac{1}{n} & 0 & & \cdots & 0 \end{bmatrix},$$

a finite-order Hankel matrix with all zeros below the main cross-diagonal. Thus "by analogy," is there a corresponding relative of the infinite Hilbert matrix A? *Answer*: A itself (plausible but not convincing?); in other words, the infinite-dimensional counterpart of Z_n ought to be A (still controversial?).

Anyhow, there is an interesting connection between the equalities $\pi = 2$ Arcsin 1 and $\|A\| = \lim_{n\to\infty}\|Z_n\|$ (instead of $\|A\| = \lim_{n\to\infty}\|A_n\|$).

PROBLEM: (1) *Show that*

$$\pi = \lim_{k\to\infty} \left\{(1k)^{-1/2} + (2(k-1))^{-1/2} + \cdots + (k1)^{-1/2}\right\}.$$

(2) *Use* (1) *or some others to show that* $\lim_{n\to\infty}\|Z_n\| \geqslant \pi$.

The combination of the results of Problems in Sections VIII and IX brings home the delightful fact $\|A\| = \pi$.

X. Two Natural Dilations. The Hilbert matrix A is a "one-way-infinite Hankel matrix." In contrast, there is a family of "two-way-infinite Hankel matrices" of the form

$$\left[\begin{array}{cc|cc} & \vdots & & \cdot^{\cdot^{\cdot}} \\ & \alpha_{-2} & \alpha_{-1} & \alpha_0 \\ \alpha_{-2} & \alpha_{-1} & \alpha_0 & \alpha_1 \\ \hline \alpha_{-1} & \alpha_0 & \alpha_1 & \alpha_2 \\ \alpha_0 & \alpha_1 & \alpha_2 & \\ \cdot^{\cdot^{\cdot}} & & \vdots & \end{array}\right].$$

Two of these matrices arise as the natural dilations of A: the first is denoted by $A^{\#}$ with

$$\alpha_k = \begin{cases} 0 & \text{if } k = 0, \\ 1/k & \text{if } k \text{ is a nonzero integer;} \end{cases}$$

the second is denoted by $A^{\dagger}$ with

$$\alpha_k = \begin{cases} 0 & \text{if } k \text{ is a nonpositive integer,} \\ 1/k & \text{if } k \text{ is a positive integer.} \end{cases}$$

PROBLEM: (1) *Show that* $\|A^{\#}\| = \pi$.
(2) *Show that* $\|A^{\dagger}\| = \infty$.

By (1), it follows again that $\|A\| \leqslant \|A^{\#}\| = \pi$. Apparently, this is the simplest way to show $\|A\| \leqslant \pi$. (Cf. problems in Sections V and VIII.)

The combination of (1) and (2) yields an extremely amazing fact: Although $A^{\dagger}$ is a lower-right-triangular part of $A^{\#}$, it turns out that $A^{\dagger}$ blows up while $A^{\#}$ is still bounded.

Alternatively, rewrite

$$A^{\#} = \begin{bmatrix} -A & B^* \\ B & A \end{bmatrix}, \quad A^{\dagger} = \begin{bmatrix} 0 & B_0^* \\ B_0 & A \end{bmatrix}$$

where B and B_0 are one-way-infinite Toeplitz matrices

$$B = \begin{bmatrix} 0 & -1 & -\frac{1}{2} & -\frac{1}{3} & \cdots \\ 1 & 0 & -1 & -\frac{1}{2} & \\ \frac{1}{2} & 1 & 0 & -1 & \\ \frac{1}{3} & \frac{1}{2} & 1 & 0 & \\ \vdots & & & & \ddots \end{bmatrix}, \quad B_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \\ \frac{1}{2} & 1 & 0 & 0 & \\ \frac{1}{3} & \frac{1}{2} & 1 & 0 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

Then it follows that $\|B\| \leqslant \|A^{\#}\| < \infty$ and $\|B_0\| = \|A^{\dagger}\| = \infty$, yet B_0 is the lower-left-triangular part of B.

SOLUTIONS AND NOTES

I. Solution. Let

$$\alpha_{ij} = (-1)^{i+j}(i+j-1)\binom{n+i-1}{n-j}\binom{n+j-1}{n-i}\binom{i+j-2}{i-1}^2.$$

The direct but cumbersome computation leading to $[\alpha_{ij}] = A_n^{-1}$ can be justified by means of mathematical inductions (on i, on j, or on n) or some general counting principles.

Alternatively, it may be easier to apply a known fact:

$$\det\left[1/(x_i + y_j)\right] = \Pi_{j>i}\left((x_j - x_i)(y_j - y_i)\right)/\Pi_{i,j}(x_i + y_j)$$

(see [**6**, p. 92, Problem 3]). Let $x_i = i - 1, y_j = j$; it follows immediately that

$$\det A_n = \frac{\{1!2! \cdots (n-1)!\}^4}{1!2! \cdots (2n-1)!}.$$

Similarly, all cofactors of A_n can be computed out explicitly. Henceforth $A_n^{-1} = [\alpha_{ij}]$ follows.

Note. The answer to this problem has been known for many years. The solution provided above has already appeared in [**9**]. See [**13**] for other related computational problems.

II. Solution. (1) Some observations are needed so as to reduce tedious computation.

(a) If A annihilates a sequence $(0,\ldots,0,1,\gamma_1,\gamma_2,\ldots)$ whose leading nonzero entry is 1, then $1 < |\gamma_1| + |\gamma_2| + \cdots$.

(b) If A annihilates a sequence $(\alpha_1, \alpha_2, \alpha_3, \ldots)$, then A also annihilates $(0, \alpha_1, \alpha_2/2, \alpha_3/3, \ldots)$.

(c) If A annihilates a nontrivial sequence, then A also annihilates a nontrivial l^1-sequence.

(d) If A annihilates a sequence $(0,\ldots,0,1,\beta_1,\beta_2,\ldots)$ whose leading nonzero entry is 1 at the pth position, then A also annihilates a sequence $(0,\ldots,0,0,1,\gamma_1,\gamma_2,\ldots)$ whose leading nonzero entry is 1 at the $(p+1)$st position and $|\gamma_j| \leqslant p|\beta_j|/(p+1)$ for all j.

Statement (a) is obvious. Statement (b) follows from the computation

$$0 = \sum_{j=1}^{\infty} \alpha_j/(i+j-1) \quad \text{for each } i = 1,2,3,\ldots$$

$$\Rightarrow 0 = \frac{1}{i}\left[\sum_{j=1}^{\infty} \alpha_j/j - \sum_{j=1}^{\infty} \alpha_j/(i+j)\right]$$

$$= \sum_{j=1}^{\infty} \alpha_j/(j(i+j)) = \sum_{j=2}^{\infty} \alpha_{j-1}/((j-1)(i+j-1)).$$

To get (c), apply (b) three times; thus if A annihilates a nontrivial sequence $(\alpha_1, \alpha_2, \alpha_3, \dots)$, then A also annihilates

$$(0,0,0,\alpha_1/(1\cdot 2\cdot 3),\dots,\alpha_j/(j(j+1)(j+2)),\dots)$$

which is an l^1-sequence in virtue of the fact $\Sigma\alpha_j/j = 0$ (from the assumption). Statement (d) is an immediate consequence of (b).

Now suppose that A is not formally one-to-one. Then by (c), we may assume that A annihilates a sequence $(0,\dots,0,1,\beta_1,\beta_2,\dots)$ whose leading coefficient is 1 at the pth position, and $|\beta_1| + |\beta_2| + \cdots < \infty$. Application of Statement (d) n times yields that A also annihilates a sequence $(0,\dots,0,1,\gamma_1,\gamma_2,\dots)$ whose leading entry is 1 at the $(p+n)$th position and $|\gamma_j| \leqslant p|\beta_j|/(p+n)$. Thus

$$|\gamma_1| + |\gamma_2| + \cdots \leqslant p(|\beta_1| + |\beta_2| + \cdots)/(p+n) \leqslant 1$$

when n is sufficiently large. But by (a), this leads to contradiction. Therefore A must be formally one-to-one.

(2) and (3). Observe that whenever A maps a sequence $(\alpha_1, \alpha_2, \alpha_3, \dots)$ to $(1,0,0,\dots)$, then A will automatically annihilate $(0, \alpha_1, \alpha_2, \alpha_3, \dots)$. Since A is formally one-to-one, it follows that the sequence $(1,0,0,\dots)$ is not in the formal range of A. This, of course, shows that A does not have a formal inverse.

Note: It is well known that $A\colon l^2 \to l^2$ is one-to-one. (See e.g. [**5**, Theorem 1, pp. 703–4] for the usual proof.) On the other hand, M. Rosenblum [**8**] has shown that for each complex number λ with $\operatorname{Re}\lambda > 0$, there is a sequence $v = (\alpha_n)$ with $\Sigma|\alpha_n|^2 = \infty$ satisfying the formal equality $Av = \lambda v$.

III. Solution. If x and r are real numbers and $r > 0, |x| < 1$, then by the binomial expansion,

$$(1-x)^{-r} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots$$

where $\alpha_0 = 1$, $\alpha_1 = r$, and each α_k is a positive real number. Thus

$$i + j - 1 = ij(1-x) \quad \text{with} \quad x = (i-1)(j-1)/(ij) < 1,$$

and

$$(i+j-1)^{-r} = i^{-r}j^{-r}(1-x)^{-r} = \sum_k \alpha_k (i-1)^k (j-1)^k/(ij)^{r+k}.$$

Notice that $[\alpha_k(i-1)^k(j-1)^k/(ij)^{r+k}]_{i,j}$ is a rank 1 positive matrix. Therefore $A_n^{[r]} = [(i+j-1)^{-r}]_{i,j}$, as a sum of positive matrices, is also positive.

Note: In particular, let $r = 1$, then $\alpha_k = 1$ for each k; thus the paragraph above provides a simple "discrete" proof for the well-known fact $A_n \geqslant 0$ (see e.g., [**1**, Section (8)] for the usual proof).

IV. Solution. As indicated in the text, the construction of B by the mathematical induction is admissible. Thus $A = BB^*$ where $B = [\beta_{ij}]$ with

$$\beta_{ij} = \begin{cases} \dfrac{\sqrt{2j-1}\,((i-1)!)^2}{(i+j-1)!(i-j)!} & \text{if } i \geqslant j, \\ 0 & \text{if } i < j. \end{cases}$$

Notes. (1) The matrix B can also be constructed as follows. Regard l^2 as $L^2[0,1]$ with the fixed orthonormal basis $\{f_1, f_2, \dots\}$ where $f_n(x)$ is a polynomial of degree $n-1$ carrying a positive coefficient in the x^{n-1} term. Let T: $L^2[0,1] \to L^2[0,1]$ be the continuous linear map sending each f_n to x^{n-1}. Henceforth, it can be verified that B^* is just the matrix representation for T.

(2) For a graceful expression, it may be better to write $A = EXDX^*E$ where

$$E = \begin{bmatrix} 1 & & & & \\ & \frac{1}{2} & & & \\ & & \frac{1}{3} & & \\ & & & \frac{1}{4} & \\ & & & & \ddots \end{bmatrix}, \quad D = \begin{bmatrix} 1 & & & & \\ & 3 & & & \\ & & 5 & & \\ & & & 7 & \\ & & & & \ddots \end{bmatrix},$$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1/3 & 0 & 0 & \\ 1 & 2/4 & 1\cdot 2/4\cdot 5 & 0 & \\ 1 & 3/5 & 2\cdot 3/5\cdot 6 & 1\cdot 2\cdot 3/5\cdot 6\cdot 7 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

Although X and D are unbounded operators: $l^2 \to l^2$, there is, however, no difficulty to carry out the formal matrix computation $B = EXD^{1/2}$, $A = BB^*$.

(3) B has a formal inverse $[\gamma_{ij}]$ where

$$\gamma_{ij} = \begin{cases} (-1)^{i+j}\sqrt{2i-1}\dbinom{i+j-2}{j-1}\dbinom{i-1}{j-1} & \text{if } i \geqslant j, \\ 0 & \text{if } i < j. \end{cases}$$

This result has been noticed by J. Todd [**12**].

V. Solution. (1) is plain since $I - (I-C)(I-C^*)$ is a diagonal operator with positive diagonal entries $1, 1/2, 1/3, \dots$.

(2) Here the equality $\|TT^*\| = \|T\|^2$ for the Hilbert-space-operator norm will be used twice. First from (1),

$$\|I - C\|^2 = \|(I-C)(I-C)^*\| \leqslant 1;$$

i.e., $\|I - C\| \leqslant 1, \|C\| \leqslant 2$. Next,

$$\|L\| = \|CC^*\| = \|C\|^2 \leqslant 4.$$

Since each entry of A is a positive real number smaller than or equal to the corresponding entry of L, it follows that $\|A\| \leqslant \|L\| \leqslant 4$ as desired.

Note: This very simple proof that $\|C\| \leqslant 2$ has appeared in [**2**, pp. 128–129].

VI. Solution. Let C be the infinite Cesáro matrix as mentioned in Section V. It is easy to see that both C: $l^2 \to l^2$ and C^*: $l^2 \to l^2$ are one-to-one (thus, C is of dense range). Next, a direct computation yields $CA = A'C$. The rest of the proof is the restatement of the following lemma which is known to many operator-theorists.

LEMMA: *Suppose T, S, and C are bounded operators satisfying $CT = SC$. If both T and S are hermitian and if C is one-to-one and of dense range, then T and S are unitarily equivalent.*

Proof. By the polar decomposition, there is a unitary U and a positive P such that $C = UP$ and P is one-to-one and of dense range (see e.g. [**3**, p. 169]). Thus

$$P^2T = PU^*UPT = C^*CT = C^*SC = (SC)^*C = (CT)^*C = TC^*C = TPU^*UP = TP^2,$$

and consequently $PT = TP$. Therefore,

$$UTP = UPT = CT = SC = SUP.$$

Since P is one-to-one and of dense range, it follows $UT = SU$; i.e., $T = U^*SU$ as desired.

VII. Solution. (1) Clearly, $A - A_n$ contains an $n \times n$ submatrix

$$Y = \begin{bmatrix} \frac{1}{2n+1} & \frac{1}{2n+2} & \cdots & \frac{1}{3n} \\ \frac{1}{2n+2} & & & \\ \vdots & & & \vdots \\ \frac{1}{3n} & & \cdots & \frac{1}{4n-1} \end{bmatrix}.$$

Let Q be the $n \times n$ matrix where every entry is $1/n$; then Q is an orthogonal projection (i.e., $Q = Q^* = Q^2$) and thus of norm 1. Since Y is entrywise larger than $Q/4$, it follows $\|A - A_n\| \geqslant \|Y\| \geqslant \|Q/4\| = 1/4$. This proves that A is not compact.

(2) First observe that $A(q) \geqslant 0$. This is an immediate consequence of the fact that A (or A') is a dilation of $A(q)$ if q is an even (or odd) integer. Next, use the notion of *Schur product*: if $T = [\tau_{ij}]$ and $S = [\sigma_{ij}]$, then the Schur product $T * S = [\tau_{ij}\sigma_{ij}]$; it is an elementary fact that if $T \geqslant 0$ and $S \geqslant 0$ then $T * S \geqslant 0$. Henceforth,

$$\begin{aligned} A - A(q) &= \left[\frac{1}{i+j-1} - \frac{1}{q+i+j-1}\right]_{ij} \\ &= \left[\frac{q}{(i+j-1)(q+i+j-1)}\right]_{ij} = qA * A(q). \end{aligned}$$

Since $A \geqslant 0$ and $A(q) \geqslant 0$, it follows that $A * A(q) \geqslant 0$ and thus $A - A(q) \geqslant 0$. Finally, observe that the sum of diagonal entries of the positive operator $A - A(q)$ is

$$\sum_j q/((2j-1)(q+2j-1)) < \infty.$$

Therefore $A - A(q)$ is a trace-class operator.

Note. It is well known (see e.g. [**1**, Section 8]) that A is not compact. It is even easier to see that $A - A(q)$ is a *Hilbert-Schmidt* operator (i.e., the sum of squares of entries of $A - A(q)$ is finite), as noted in [**1**, Section 8].

VIII. Solution. First define an auxiliary function Φ: $\mathbb{Z}^2 \to \mathbb{R}$ by

$$\Phi(k,l) = \begin{cases} 1/k & \text{if } k \neq 0, l = 0 \\ 1/l & \text{if } k = 0, l \neq 0 \\ -1/k & \text{if } k = l \neq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

It follows that

(a) Φ, regarded as an ordered set, is an Euler vector,

(b) if (k_0, l_0) is a fixed pair of integers different from $(0,0)$, then

$$\sum_{k,l} \Phi(k,l)\Phi(k+k_0, l+l_0) = 0.$$

Next, let $\mathcal{H}_0$ be the Hilbert space l^2 with the orthonormal basis $\{f_1, f_2, \ldots\}$ and let $\mathcal{H}_1$ be $l^2(\mathbb{Z}^2)$ with the orthonormal basis $\{e_\mu\colon \mu \in \mathbb{Z}^2\}$; imbed $\mathcal{H}_0$ into $\mathcal{H}_1$ by identifying each f_j with e_μ

where $\mu = (j,0)$. Thus the infinite Hilbert matrix A (or A') stands for an operator: $\mathcal{H}_0 \to \mathcal{H}_0$ while two matrices $T = [\tau_{\mu\nu}]$, $S = [\sigma_{\mu\nu}]$ ($\mu, \nu \in \mathbb{Z}^2$) to be constructed below are operators: $\mathcal{H}_1 \to \mathcal{H}_1$. Now, define

$$\left.\begin{aligned}\tau_{\mu\nu} &= \Phi(k-1,l)\\ \sigma_{\mu\nu} &= \Phi(k,l)\end{aligned}\right. \quad \text{if } \mu+\nu=(k,l)\in\mathbb{Z}^2.$$

Consequently, if i and j are positive integers and $\mu = (i,0)$, $\nu = (j,0)$, then $\tau_{\mu\nu} = 1/(i+j-1)$, $\sigma_{\mu\nu} = 1/(i+j)$; i.e., T is a dilation of A, and S is a dilation of A'. From (a) and (b), it follows that each column vector of T—i.e., $(\tau_{\mu\nu})_\mu$ for each fixed ν—is an Euler vector and any two column vectors of T are orthogonal (the similar result holds for S).

Note. Alternatively, the operators T and S in the solution above can be constructed as follows. Let $\mathbb{T}$ be the unit circle with the normalized Lebesgue measure μ, let $\mathcal{H} = L^2(\mathbb{T}\times\mathbb{T}, \mu\times\mu)$, and let M: $\mathcal{H}\to\mathcal{H}$ be an operator defined by

$$g(z,w) \overset{M}{\mapsto} h(z,w) = \begin{cases} i\pi g(\bar z,\bar w) & \text{if } \operatorname{Arg} z + \operatorname{Arg} w \leqslant 2\pi,\\ -i\pi g(\bar z,\bar w) & \text{if } \operatorname{Arg} z + \operatorname{Arg} w > 2\pi,\end{cases}$$

subject to the constraint $2\pi > \operatorname{Arg} z \geqslant 0, 2\pi > \operatorname{Arg} w \geqslant 0$. Then it follows immediately that $M = M^*$, $M^2 = \pi^2 I$. Moreover, there is a natural way (respectively, a quasi-natural way) to identify $l^2(\mathbb{Z}^2)$ with $L^2(\mathbb{T}\times\mathbb{T})$: let $e_\mu = z^k w^l$ (resp., $e_\mu = z^{k-1}w^l$) whenever $\mu = (k,l)\in\mathbb{Z}^2$. Henceforth, it can be verified that M has the matrix expression $[\sigma_{\mu\nu}]$ (resp., $[\tau_{\mu\nu}]$) as in the solution above.

IX. Solution. (1) $\Sigma_{j=0}^{k-1}(j(k-j))^{-1/2} = \frac{1}{k}\Sigma_{j=0}^{k-1}\left(\frac{j}{k}\left(1-\frac{j}{k}\right)\right)^{-1/2}$ is a Riemann sum for the function $(x(1-x))^{-1/2}$ on $(0,1)$.

$$\therefore \lim_{k\to\infty}\sum_{j=0}^{k-1}(j(k-j))^{-1/2} = \int_0^1 (x(1-x))^{-1/2} = 2\operatorname{Arcsin} 1 = \pi.$$

(2) Let v be the column vector $(1, 2^{-1/2}, \dots n^{-1/2})$. Then

$$\|Z_n\| \geqslant (Z_n v, v)/\|v\|^2 = \left(\sum_{k=1}^n a_k/k\right)\Big/\left(\sum_{k=1}^n 1/k\right)$$

where $a_k = \Sigma_{j=0}^{k-1}(j(k-j))^{-1/2}$. Since $\lim_{k\to\infty} a_n = \pi$ (from (1)), it follows by general computation $\lim_{n\to\infty}(\Sigma_{k=1}^n a_k/k)/(\Sigma_{k=1}^n 1/k) = \lim_{k\to\infty} a_k = \pi$; thus $\lim_{n\to\infty}\|Z_n\| \geqslant \pi$ as desired.

Note. The main ingredient of the proof above has appeared in [**11**]. There are other known proofs of $\|A\| = \pi$ in the literature (see e.g. [**4**, Chapter 9 and Appendix III] [**10**, Chapter 9, p. 101]).

X. Lemma. *Let T and H be two-way-infinite matrices of the form*

$$T = \left[\begin{array}{cccc|cccc} \ddots & & & \vdots & & & & \\ & & \alpha_0 & \alpha_1 & \alpha_2 & & & \\ & & \alpha_{-1} & \alpha_0 & \alpha_1 & \alpha_2 & & \\ \hline & & \alpha_{-2} & \alpha_{-1} & \alpha_0 & \alpha_1 & & \\ & & & \alpha_{-2} & \alpha_{-1} & \alpha_0 & & \\ & & & & \vdots & & \ddots & \end{array}\right], \quad H = \left[\begin{array}{cccc|cccc} & & & \vdots & & & \iddots & \\ & & & \alpha_{-2} & \alpha_{-1} & \alpha_0 & & \\ & & \alpha_{-2} & \alpha_{-1} & \alpha_0 & \alpha_1 & & \\ \hline & & \alpha_{-1} & \alpha_0 & \alpha_1 & \alpha_2 & & \\ & & \alpha_0 & \alpha_1 & \alpha_2 & & & \\ & \iddots & & & \vdots & & & \end{array}\right].$$

If $\alpha_k \geqslant 0$ for each integer k, then $\|T\| = \|H\| = \Sigma_{-\infty}^{\infty}\alpha_k$.

Proof. Let S_k be the matrix in the same form as T with ones along the kth diagonal and zeros elsewhere. Then it is obvious that $\|S_k\| = 1$ and $\|T\| = \|\Sigma\alpha_k S_k\| \leqslant \Sigma\alpha_k$. On the other hand, consider the $n \times n$ Toeplitz matrix

$$X = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_{-1} & \alpha_0 & \alpha_1 & & \\ \alpha_{-2} & \alpha_{-1} & \alpha_0 & & \\ \vdots & & & \ddots & \\ \alpha_{-n} & & & & \alpha_0 \end{bmatrix}$$

(as a submatrix of T) and the constant column vector $v = (1, 1, \dots, 1) \in \mathbb{C}^n$. Then

$$\begin{aligned} \|T\| &\geqslant (Xv, v)/\|v\|^2 \\ &= \alpha_0 + \left(1 - \frac{1}{n}\right)(\alpha_1 + \alpha_{-1}) + \left(1 - \frac{2}{n}\right)(\alpha_2 + \alpha_{-2}) + \cdots + \frac{1}{n}(\alpha_n + \alpha_{-n}). \end{aligned}$$

When n approaches to infinity, the right-hand side approaches to $\Sigma\alpha_k$. Thus $\|T\| \geqslant \Sigma\alpha_k$. Therefore the equality $\|T\| = \Sigma\alpha_k$ is verified.

The proof above can be modified so as to yield the similar result for H. Alternatively, let U be the unitary matrix in the same form as H with ones along the main cross-diagonal and zeros everywhere; then it follows that $T = HU$, and thus $\|T\| = \|H\|$ as desired.

Solution. (1) Euler's formula will be used twice here. First by direct computation, $(A^{\#})^2$ is in the same form as T in the Lemma with $\alpha_0 = \pi^2/3$ and $\alpha_k = 2/k^2$ if k is a nonzero integer. Since $A^{\#}$ is hermitian, it follows that

$$\|A^{\#}\|^2 = \|(A^{\#})^2\| = \sum_{-\infty}^{\infty} \alpha_k = \pi^2.$$

Therefore $\|A^{\#}\| = \pi$ as desired.

(2) follows from the Lemma.

Note. The fact $\|B\| < \pi$ and $\|B_0\| = \infty$ is well known (see [**4**, § 9.6]).

Acknowledgements. This research was supported in part by NSERC of Canada. The author would like to express his thanks to Professors Paul Halmos, John Todd, and Olga Taussky Todd for stimulating communications on the Hilbert matrix.

References

1. J. Barrıa and P. R. Halmos, Asymptotic Toeplitz operators, Trans. Amer. Math. Soc., 273 (1982)621–630.

2. A. Brown, P. R. Halmos, and A. L. Shields, Cesaro operators, Acta Szeged, 26 (1965)125–137.

3. P. R. Halmos, A Hilbert Space Problem Book, Van Nostrand, Princeton, 1967.

4. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, Cambridge, 1964.

5. W. Magnus, On the spectrum of Hilbert's matrix, Amer. J. Math., 72 (1950)699–704.

6. G. Pólya and G. Szegő, Problems and Theorems in Analysis, vol. 2, Springer-Verlag, Heidelberg, 1976.

7. S. Power, The essential spectrum of a Hankel operator with piecewise continuous symbol, Michigan Math. J., 25 (1978)117–121.

8. M. Rosenblum, On the Hilbert matrix, I, Proc. Amer. Math. Soc., 9 (1958)137–140.

9. I. R. Savage and E. Lukacs, Tables of inverses of finite segments of the Hilbert matrix, in Contributions to the Solutions of Systems of Linear Equations and the Determination of Eigenvalues, edited by O. Taussky, National Bureau of Standards Applied Mathematics Series, 39, 1954, 105–108.

10. D. Sarason, Function Theory on the Unit Circle, Lecture notes for a conference at Virginia Polytechnic and State University, 1978.

11. O. Taussky, A remark concerning the characteristic roots of the finite segments of the Hilbert matrix, Quart. J. Math. Oxford Ser. (2), (1949)80–83.

12. J. Todd, The conditions of the finite segments of the Hilbert matrix, in Contributions to the Solutions of Systems of Linear Equations and the Determination of Eigenvalues, edited by O. Taussky, National Bureau of Standards Applied Mathematics Series, 39, 1954, 109–116.

13. J. Todd, Computational problems concerning the Hilbert matrix, J. Res. Nat. Bur. Standards Sect. B, Math. and Math. Physics, 65 (1961)19–22.

A GROUP WITH CONSTANT GROWTH RATE

R. HIRSHON

Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201

1. Introduction. The theory of direct products in groups and related constructions in algebra is a rich and varied field. A sample list of some works (by no means complete) is included in the references. Problems often center on obtaining certain invariants for the decompositions involved. Sometimes the problems that arise are easy to state and comprehend. However, as in elementary number theory, some of these problems await fresh insights and new methods for their solutions. Sometimes the results obtained are somewhat anomalous insofar as they show that direct products do not always behave as one might expect. It is only after looking at this construction for some time that one begins to accept a general principle. This is: Almost anything can happen! In this paper we will discuss a result which seems to violate the intuition of our formative years.

2. A Review of Some Basics. Before stating the exact problem to be discussed, a brief review of the fundamental concepts which occur might be useful. If A and B are isomorphic groups, we write $A \approx B$.

2.1 *Direct Products.* We say that a group G is the direct products of its n subgroups

$$G_1, G_2, \ldots, G_n$$

if three conditions hold. In the first place, every element g in G must be expressible as a product

$$g = g_1 \cdot g_2 \cdots g_n, \quad g_i \in G_i. \tag{1}$$

Secondly, if $w \in G_i$ and $v \in G_j$ and $i \neq j$ then $wv = vw$. Finally the expression in (1) must be unique. That is, if in addition to (1) we also have

$$g = \bar{g}_1 \cdot \bar{g}_2 \cdots \bar{g}_n, \quad \bar{g}_i \in G_i$$

then $\bar{g}_i = g_i$ for all i. When these three conditions hold we write

$$G = G_1 \times G_2 \times \cdots \times G_n.$$

The above definition (internal definition) starts with the group G. A natural question which arises is the following. If one is given n arbitrary groups

$$G_1, G_2, \ldots, G_n \tag{2}$$

which are not necessarily distinct, is there a group G with

$$\overline{G} = \overline{G}_1 \times \overline{G}_2 \times \cdots \times \overline{G}_n$$

with $\overline{G}_i \approx G_i$ for all i? The answer is yes. Define $\overline{G}$ as the cartesian product of the G_i. That is, $\overline{G}$ is the set of all n-tuples α of the form

Ronald Hirshon received his Ph.D. in 1967 from Adelphi University under the direction of Donald Solitar. He has taught at the Polytechnic Institute of New York since 1967 where he is currently a Professor of Mathematics. He is fond of cycling, swimming, reading, the movies and his family (not necessarily in the given order). He is currently at work on a study of direct products entitled *Some J Replacement Results*.

12. J. Todd, The conditions of the finite segments of the Hilbert matrix, in Contributions to the Solutions of Systems of Linear Equations and the Determination of Eigenvalues, edited by O. Taussky, National Bureau of Standards Applied Mathematics Series, 39, 1954, 109–116.

13. J. Todd, Computational problems concerning the Hilbert matrix, J. Res. Nat. Bur. Standards Sect. B, Math. and Math. Physics, 65 (1961)19–22.

A GROUP WITH CONSTANT GROWTH RATE

R. HIRSHON

Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201

1. Introduction. The theory of direct products in groups and related constructions in algebra is a rich and varied field. A sample list of some works (by no means complete) is included in the references. Problems often center on obtaining certain invariants for the decompositions involved. Sometimes the problems that arise are easy to state and comprehend. However, as in elementary number theory, some of these problems await fresh insights and new methods for their solutions. Sometimes the results obtained are somewhat anomalous insofar as they show that direct products do not always behave as one might expect. It is only after looking at this construction for some time that one begins to accept a general principle. This is: Almost anything can happen! In this paper we will discuss a result which seems to violate the intuition of our formative years.

2. A Review of Some Basics. Before stating the exact problem to be discussed, a brief review of the fundamental concepts which occur might be useful. If A and B are isomorphic groups, we write $A \approx B$.

2.1 *Direct Products.* We say that a group G is the direct products of its n subgroups

$$G_1, G_2, \ldots, G_n$$

if three conditions hold. In the first place, every element g in G must be expressible as a product

$$g = g_1 \cdot g_2 \cdots g_n, \quad g_i \in G_i. \tag{1}$$

Secondly, if $w \in G_i$ and $v \in G_j$ and $i \neq j$ then $wv = vw$. Finally the expression in (1) must be unique. That is, if in addition to (1) we also have

$$g = \bar{g}_1 \cdot \bar{g}_2 \cdots \bar{g}_n, \quad \bar{g}_i \in G_i$$

then $\bar{g}_i = g_i$ for all i. When these three conditions hold we write

$$G = G_1 \times G_2 \times \cdots \times G_n.$$

The above definition (internal definition) starts with the group G. A natural question which arises is the following. If one is given n arbitrary groups

$$G_1, G_2, \ldots, G_n \tag{2}$$

which are not necessarily distinct, is there a group G with

$$\overline{G} = \overline{G}_1 \times \overline{G}_2 \times \cdots \times \overline{G}_n$$

with $\overline{G}_i \approx G_i$ for all i? The answer is yes. Define $\overline{G}$ as the cartesian product of the G_i. That is, $\overline{G}$ is the set of all n-tuples α of the form

Ronald Hirshon received his Ph.D. in 1967 from Adelphi University under the direction of Donald Solitar. He has taught at the Polytechnic Institute of New York since 1967 where he is currently a Professor of Mathematics. He is fond of cycling, swimming, reading, the movies and his family (not necessarily in the given order). He is currently at work on a study of direct products entitled *Some J Replacement Results*.

$$\alpha = (g_1, g_2, \ldots, g_n), \quad g_i \in G_i.$$

We say that g_i is the ith coordinate of α. If $\beta = (h_1, h_2, \ldots, h_n)$ is in $\overline{G}$, $\alpha\beta$ is defined by

$$\alpha\beta = (g_1h_1, g_2h_2, \ldots, g_nh_n).$$

It is easy to verify that $\overline{G}$ is a group under multiplication of n-tuples. Furthermore if we define $\overline{G}_i$ as the set of all n-tuples of $\overline{G}$ whose jth coordinate for all j, $j \neq i$ is the identity of G_j, then $\overline{G}_i \approx G_i$ and $\overline{G} = \overline{G}_1 \times \overline{G}_2 \times \overline{G}_3 \times \cdots \times \overline{G}_n$. We will be especially concerned with this construction when all the groups in (2) are indentical. That is

$$G_1 = G_2 = G_3 = \cdots = G_n = H.$$

In this case we call the resulting group (or any group isomorphic to the resulting group) the direct product of n copies of H and write $\overline{G} = H^n$ for short. In dealing with direct products and in particular with H^n, the construction of $\overline{G}$ shows that by considering the groups $\overline{G}_i$ we may always use the internal definition of the direct product.

2.2 *Finitely Generated Groups.* A group H is said to be finitely generated if H contains a finite set of r elements $h_1, h_2, \ldots, h_r$ such that every element h in H may be expressed in at least one way as a product of one or more of the elements

$$h_1, h_2, \ldots, h_r, h_1^{-1}, h_2^{-1}, \ldots, h_r^{-1}.$$

In this case we say that H is generated by $h_1, h_2, \ldots, h_r$ or that $h_1, h_2, \ldots, h_r$ are a set of generators of H. Clearly every finite group is finitely generated. Standard arguments on cardinality show that a finitely generated group is countable so that for example the group of real numbers under addition is not finitely generated. If H is a finitely generated group let $d(H)$ designate the least positive integer r such that H is generated by r of its elements. For example $d(H) = 1$ if and only if H is a cyclic group. On the other hand if H is the symmetric group on n elements $d(H) = 2$ ([**44**], v.1, p. 49). If H is the infinite cyclic group, one can easily show $d(H^n) = n$.

2.3 *Finitely Presented Groups.* Let us begin with an example. Let G be the symmetric group on three elements 1, 2, 3. Let α and β be the cyclic permutations $\alpha = (1,2)$ $\beta = (1,2,3)$. Then α and β generate G and

$$\alpha^{-1}\beta\alpha = \beta^2, \quad \alpha^2 = 1, \quad \beta^3 = 1. \tag{3}$$

It is possible to show that any other relation involving α and β is a consequence of the three relations (3) and the trivial relations $yy^{-1} = 1$ where y is α, β, α^{-1} or β^{-1}. One describes this situation by saying G is generated by the elements α and β subject to the defining relations (3). The elements α, β together with the relations (3) are said to be a presentation of G. For short we write

$$G = \langle \alpha, \beta; \quad \alpha^{-1}\beta\alpha = \beta^2, \quad \alpha^2 = \beta^3 = 1\rangle \tag{4}$$

and refer to (4) as a presentation of G. One of the nice results of presentation theory tells us that anything we write down that looks like a presentation in fact turns out to be the presentation of some group. For example if we write four symbols a, b, c, d and if we write

$$b^{-1}ab = a^2, \quad c^{-1}bc = b^2, \quad d^{-1}cd = d^2, \quad a^{-1}da = d^2 \tag{5}$$

and designate these relations by P_1, P_2, P_3, P_4, then there is exactly one group (up to an isomorphism) that has the presentation

$$\langle a, b, c, d; \quad P_1, P_2, P_3, P_4\rangle. \tag{6}$$

A finitely presented group is one which has a presentation involving a finite number of generators and a finite number of relations among these generators. Of course the presentation that one writes down may turn out to define the identity group. For example the group on one generator w

subject to the relations $w^2 = 1$ and $w^3 = 1$ must clearly define the identity group. A little less obvious is the fact that the group on three generators a, b, c subject to the relations

$$b^{-1}ab = a^2, \quad c^{-1}bc = b^2, \quad a^{-1}ca = c^2$$

defines the trivial group [**18**]. However the group (6) turns out to be an infinite group [**18**]. The problem of determining whether or not a given presentation defines the trivial group is often very difficult. In fact it has been shown that it is impossible to describe a general method which works in a finite number of steps for all presentations [**50**]. A very clear exposition of the elements of presentation theory can be found in the first few chapters of [**46**].

3. Groups with Constant Growth Rate. One seems to feel that if G is a finitely generated nontrivial group that

$$\lim_{n\to\infty} d(G^n) = \infty. \tag{7}$$

Indeed if G is a nontrivial finite group, (7) is true [**12**], [**60**]. Using the fact that every finitely generated abelian group is a direct product of cyclic groups [**44**] one can easily see that (7) is true if G is a finitely generated abelian group. If G and G_* are finitely generated and G_* is a homomorphic image of G, then $d(G_*) \leqslant d(G)$. In this case G_*^n is also a homomorphic image of G^n so $d(G_*^n) \leqslant d(G^n)$. Consequently one sees that if G is a finitely generated group and G has a nontrivial finite homomorphic image, then again (7) is true. An interesting study of the behavior of $d(G^n)$ is made in [**60**]. For example, the growth rate of $d(G^n)$ for G a finite perfect group is quite slow. If A_5 is the alternating group of five elements, $A_5^{6,464,040}$ requires only five generators [**60**].

At the other end of the spectrum, M. Jones has discovered an amazing nontrivial finitely generated group G such that $G^2 \approx G$ [**38**].* Hence Jones' group has the property that $G^n \approx G$ and $d(G^n) = d(G)$ for all n.

The construction of Jones' group is fairly complicated. The purpose of this note is to show that the infinite group H which is given by (6) has the property that $d(H^n) = 4$ for all n. It seems then that H might be one of the simplest groups whose growth rate $d(H^n)$ is constant.

4. Proof that $d(H^n) = 4$. To begin with it is useful to note that (5) implies

$$b^{-r}a^sb^r = a^{2^rs}, b^ra^{s2^j}b^{-r} = a^{s2^{j-r}}, \quad j \geqslant r \geqslant 0$$

and corresponding relations for the other generators. We assume that H^n is the direct product of its subgroups $H_1, H_2, \ldots, H_n$ where H_i is generated by a_i, b_i, c_i, d_i subject to the relations

$$b_i^{-1}a_ib_i = a_i^2, \quad c_i^{-1}b_ic_i = b_i^2, \quad d_i^{-1}c_id_i = c_i^2, \quad a_i^{-1}d_ia_i = d_i^2. \tag{8}$$

Let

$$\begin{aligned} a_1a_2^2a_3^{2^2} \cdots a_j^{2^{j-1}} &= w_1(j) \\ b_1b_2^2b_3^{2^2} \cdots b_j^{2^{j-1}} &= w_2(j) \\ c_1c_2^2c_3^{2^2} \cdots c_j^{2^{j-1}} &= w_3(j) \\ d_1d_2^2d_3^{2^2} \cdots d_j^{2^{j-1}} &= w_4(j) \end{aligned}$$

and let

$$w_1 = w_1(n), \quad w_2 = w_2(n), \quad w_3 = w_3(n), \quad w_4 = w_4(n).$$

We claim that H^n is generated by w_1, w_2, w_3, w_4. To outline the idea of the proof, let L be the

*The temptation at this point to cite one of those easily understood but perhaps difficult problems is irresistible. In the 1950's Peter Hilton asked whether one could find a nontrivial finitely presented group with $G^2 \approx G$. The problem is unsolved today.

subgroup of H^n generated by w_1, w_2, w_3, w_4. Note first because of the symmetry in (8), for any fact concerning a_n, b_n, c_n, d_n there is a corresponding fact obtainable by permuting these generators. Now note that it suffices to show

$$L \text{ contains } a_n, b_n, c_n \text{ and } d_n. \tag{9}$$

For if we can show (9), then it follows that if L_j is the subgroup of L generated by $w_1(j), w_2(j), w_3(j), w_4(j)$ then

$$L = L_{n-1} \times H_n. \tag{10}$$

However any proof that shows (9) will at the same time yield a proof that shows

$$L_{n-1} \text{ contains } a_{n-1}, b_{n-1}, c_{n-1}, d_{n-1} \tag{11}$$

so that $L_{n-1} = L_{n-2} \times H_{n-1}$. Clearly, by repeating this inductive procedure we obtain our assertion.

Now to show (9) it suffices to show

$$L \text{ contains } a_n \tag{12}$$

because by symmetry, any proof that shows L contains a_n will work to show L contains b_n, c_n and d_n. In order to show (12), it suffices to show that

$$L \text{ contains an odd power of } a_n \tag{13}$$

and L contains an element of the form

$$z = ya_n, \quad y \in H_1 \times H_2 \times \cdots \times H_{n-1}. \tag{14}$$

For if we can show (13), then by symmetry, L contains an odd power of b_n, c_n and of d_n. Let α_1, α_2, α_3 and α_4 be the least positive powers of a_n, b_n, c_n and d_n which are in L. By symmetry, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha$. By the Euclidean algorithm, if a_1^γ is in L, α is a divisor of γ. Hence if (13) is true, α is odd. But then from (14) L contains $b_n^{-\alpha} y a_n b_n^\alpha = \bar{y}$. But $\bar{y} = ya_n^{2^\alpha}$ so that $(ya_n^{2^\alpha})(ya_n)^{-1}$ is in L. Hence if $\beta = 2^\alpha - 1$, a_n^β is in L and α divides β. If α is greater than 1, this is impossible. For if $\alpha > 1$, let p be the least prime divisor of α. Then $2^\alpha \equiv 1 \bmod p$ and $2^{p-1} \equiv 1 \bmod p$. Hence if the order of 2 mod p is d, d divides α and d divides $p - 1$ contradicting the minimality of p. Hence $\alpha = 1$ so that (12) is true. We complete the proof by showing (13) and (14).

To show (13) consider all elements in L of the form,

$$v = a_k^{\theta_k} a_{k+1}^{\theta_{k+1}} \cdots a_n^{\theta_n} \tag{15}$$

where the exponent θ_n of a_n is positive. We call v an $n - k + 1$ string. For example, w_1 is an n string. If v is an $n - k + 1$ string given by (15), then an easy calculation shows that if $\delta_j = 2^{2^j}$, then

$$w_2^{-1} v w_2 v^{-\delta_{k-1}} = a_{k+1}^{\varepsilon_{k+1}} \cdots a_n^{\varepsilon_n} \tag{16}$$

where $\varepsilon_i = \theta_i(\delta_{i-1} - \delta_{k-1})$. In particular $\varepsilon_n > 0$ so that (16) yields an $n - k$ string in L. By induction, there is a 1 string in L. That is L contains an element of the form a_n^θ, $\theta > 0$. Let $\theta = 2^\delta \theta_*$ where θ_* and 2 are relatively prime. Let $\eta = 2^{n-1}$, $a = a_n$. If $\delta \leqslant \eta$, then

$$(a^\theta)^{2^{\eta-\delta}} = a^{2^\eta \theta_*}$$

so that $a^{2^\eta \theta_*}$ is in L. But then

$$w_2 a^{2^\eta \theta_*} w_2^{-1} = a^{\theta_*}$$

so a^{θ_*} is in L. If $\delta > \eta$, let $\delta = \mu\eta + \mu_0, 0 \leqslant \mu_0 < \eta$. Then

$$w_2^\mu a^\theta w_2^{-\mu} = a^{2^{\mu_0}\theta_*}$$

so that $a^{2^{\mu_0}\theta_*}$ is in L and as above this implies that a^{θ_*} is in L and (13) is shown. Finally if we choose integers μ_1 and μ_2 with $\mu_1 2^{n-1} + \mu_2\theta_* = 1$, then

$$w_1^{\mu_1}a^{\mu_2\theta_*} = w_1(n-1)^{\mu_1}a$$

so that we may take $y = w_1(n-1)^{\mu_1}$.

In concluding we pose a question which we make no attempt to answer. Does there exist a finitely generated group H_*, $H_* \neq 1$ such that H_* is defined by three or fewer than three defining relations and such that the growth rate $d(H_*^n)$ is constant?

References

1. R. Appleson and L. Lovasz, A characterization of cancellable K-ary structures, Period. Math. Hungar. G. (1975) 17–19.

2. G. Baumslag, Some metacyclic groups with the same finite images, Compositio Math., 29 (1974) 249–252.

3. ______, Direct decompositions of finitely generated torsion free nilpotent groups, Math. Z., 145 (1975) 1 – 10.

4. R. Baer, The decomposition of enumerable primary abelian groups into direct summands, Quart. J. Math. Oxford, 6 (1935) 217–221.

5. ______, Direct decompositions, Trans. Amer. Math. Soc., 62 (1947) 62–98.

6. ______, The role of the center in the theory of direct decompositions, Bull. Amer. Math. Soc., 64 (1948) 519–551.

7. C. Chang, B. Jónsson and A. Tarski, Refinement properties for relational structures, Fund. Math., 53 (1964) 249–281.

8. I. Cohen and I. Kaplansky, Rings for which every module is a direct sum of cyclic modules, Math. Z., 54 (1951) 97–101.

9. P. M. Cohn, The complement of a finitely generated direct summand of an abelian group, Proc. Amer. Math. Soc., 7 (1956) 520–521.

10. P. Crawley, The cancellation of torsion abelian groups in direct sums, J. Algebra, 2 (1965) 432–442.

11. P. Crawley and B. Jónsson, Refinements for infinite direct decompositions of algebraic systems, Pacific J. Math., 14 (1964) 797–885.

12. I. M. S. Dey, Embeddings in non-hopf groups, J. London Math. Soc., (2) 1 (1969) 745–749.

13. E. G. Evans, Jr., Krull-Schmidt and cancellation over local rings, Pacific J. Math., 46 (1973) 115–121.

14. L. Fuchs, The direct sum of cyclic groups, Acta Math. Acad. Sci. Hungar., 3 (1952) 177–195.

15. ______, On a substitution property for modules, Monatsh. Math., 75 (1971) 198–204.

16. O. N. Golowin, On factors without centers in direct decompositions of groups, Rec. (Math. Sbornik) NS, 6 (1939) 423–436.

17. K. R. Goodearl, Power cancellation of groups and modules, P. J. Math., 64 (1976) 387–411.

18. G. Higman, A finitely generated infinite simple group, J. London Math. Soc., 26 (1951) 61–64.

19. R. Hirshon, On hopfian groups, Pacific J. Math., 32 (1970) 753–766.

20. ______, Some theorems on hopficity, Trans. Amer. Math. Soc., 141 (1969) 299–344.

21. ______, Cancellation of group with maximal condition, Proc. Amer. Math. Soc., 24 (1970) 401–403.

22. ______, On cancellation in groups, this MONTHLY, 76 (1969) 1037–1039.

23. ______, The center and the commutator subgroup in hopfian groups, J. Arkiv fur Mathematatik, 9 (1971) 181–192.

24. ______, A conjecture on hopficity and related results, Arch. Math., 22 (1971) 449–455.

25. ______, New groups admitting essentially unique directly indecomposable decompositions, Math. Ann., 194 (1971) 123–125.

26. ______, A problem in group theory, this MONTHLY (Research Section), 89 (1972) 379–380.

27. ______, The direct product of a hopfian group with a group with a cyclic center, Arch. Math., 10 (1972) 231–234.

28. ______, The cancellation of an infinite cycle group in direct products, Arch. Math., 26 (1975) 134–479.

29. ______, The direct product of a hopfian group with a p group, Arch. Math., 26 (1975) 470–479.

30. ______, Some cancellation theorems with applications to nilpotent groups, J. Austral. Math. Soc., 23 (1977) 147–166.

31. ______, Cancellation and hopficity in direct products, J. Algebra, 50 (1978) 26–32.

32. ______, The equivalence of $x^tC \approx x^tD$ and $J \times C \approx J \times D$, Trans. Amer. Math. Soc., 249 (1979-A) 331–340.

33. ______, Decompositions of groups with finitely generated commutator quotient group, J. Austral. Math. Soc. Ser. A, 28 (1979-B) 315–320.
34. ______, The number of indecomposable terms in direct decompositions of groups with the maximal condition, J. Algebra, 75 (1982) 70–81.
35. ______, Some J replacement results, manuscript.
36. B. Jónsson and A. Tarski, Direct decompositions of finite algebraic systems, Notre Dame Mathematical Lectures No. 5, 1947.
37. B. Jónsson, On direct decompositions of torsion free abelian groups, Math. Scand., 7 (1959) 361–371.
38. J. M. Tyrer Jones, Direct products and the hopf property, J. Austral. Math. Soc., 27 (1974) 174–198.
39. A. Kertész, On groups every subgroup of which is a direct summand, Publ. Math. Debrecen, 2 (1951) 74–75.
40. A. Kertész, and T. Szele, On abelian groups every multiple of which is a direct summand, Acta Sci. Math. (Szeged), 14 (1952) 157–166.
41. V. Korinek, Sur la décomposition d'un groupe en produit direct des sous groupes, Časopis Pest. Mat., vol. 63 (1937) 261–286.
42. A. Kurosh, Isomorphisms of direct decompositions I, Bull. Acad. Sci. USSR, 7 (1943) 184–202.
43. ______, Isomorphisms of direct decompositions II, Bull. Acad. Sci. USSR, 10 (1946) 47–72.
44. ______, The Theory of Groups, translated by K. A. Hirsch from the second Russian edition, vol. 1, 2, Chelsea, New York, 1956.
45. L. Lovasz, On the cancellation law among finite relational structures, Periodic. Math. Hungar., 1 (1971) 145–156.
46. Magnus, Karrass, Solitar, Combinatorial Group Theory, Dover Publications, 1966.
47. R. McKenzie, A method for obtaining refinement theorems with an application to direct products of semigroups, Algebra Universalis, 2 (1972) 324–338.
48. G. S. Monk, A characterization of exchange rings, Proc. Amer. Math. Soc., 35 (1972) 349–353.
49. O. Ore, Direct decompositions, Duke Math. J., 2 (1936) 581–596.
50. M. O. Rabin, Recursive unsolvability of group theoretic problems, Ann. of Math., 67 (1958) 172–194.
51. R. Remak, Uber die Zerlegunigen der endlichen gruppen in direkte unzerlegbure factoren, J. Reine Angew. Math., 139 (1911) 293–308.
52. S. Shastri, Some cancellation theorems, J. Austral. Math. Soc., 30 (1980) 87–97.
53. W. V. Vasconcelos, On local and stable cancellation, An. Acad. Brasil. Ciênc. 37 (1965) 389–393.
54. E. A. Walker, Cancellation in direct sums of groups, Proc. Amer. Math. Soc., 7 (1956) 898–902.
55. R. B. Warfield, Jr., Genus and cancellation for groups with finite commutator subgroups, J. Pure Appl. Algebra, 6 (1975) 125–132.
56. ______, Cancellation and exchange theorems for finitely generated nilpotent groups.
57. ______, A Krull-Schmidt theorem for infinite sums of modules, Proc. Amer. Math. Soc., 22 (1969) 460–465.
58. ______, Exchange rings and decompositions of modules, Math. Ann., 199 (1972) 31–36.
59. J. H. M. Wedderburn, On the direct product in the theory of finite groups, Ann. of Math., 10 (1909) 173–176.
60. J. Wiegold, Growth of finite groups, J. Austral. Math. Soc., 27 (1974) 133–143.

CALCULATOR FUNCTION APPROXIMATION

CHARLES W. SCHELIN
Department of Mathematics, University of Wisconsin-La Crosse, La Crosse, WI 54601

1. Introduction. For about the past 10 years elementary function values have been available at the touch of a button on a hand-held calculator. The unified algorithm used in most hand calculators to approximate elementary functions is the subject of this paper. Surprisingly, this general algorithm which employs bit by bit approximation does not depend on results or

The author received his Ph.D. in 1971 from the University of Montana working in numerical analysis under Charles Bryan. Since that time he has taught at UWL, spending summers as a research fellow at McDonnell Douglas, EXXON, NASA-Ames, and Iowa State University. This paper was written while the author was an American Council on Education fellow in academic administration at the University of Utah.

33. ______, Decompositions of groups with finitely generated commutator quotient group, J. Austral. Math. Soc. Ser. A, 28 (1979-B) 315–320.
34. ______, The number of indecomposable terms in direct decompositions of groups with the maximal condition, J. Algebra, 75 (1982) 70–81.
35. ______, Some J replacement results, manuscript.
36. B. Jónsson and A. Tarski, Direct decompositions of finite algebraic systems, Notre Dame Mathematical Lectures No. 5, 1947.
37. B. Jónsson, On direct decompositions of torsion free abelian groups, Math. Scand., 7 (1959) 361–371.
38. J. M. Tyrer Jones, Direct products and the hopf property, J. Austral. Math. Soc., 27 (1974) 174–198.
39. A. Kertész, On groups every subgroup of which is a direct summand, Publ. Math. Debrecen, 2 (1951) 74–75.
40. A. Kertész, and T. Szele, On abelian groups every multiple of which is a direct summand, Acta Sci. Math. (Szeged), 14 (1952) 157–166.
41. V. Korinek, Sur la décomposition d'un groupe en produit direct des sous groupes, Časopis Pest. Mat., vol. 63 (1937) 261–286.
42. A. Kurosh, Isomorphisms of direct decompositions I, Bull. Acad. Sci. USSR, 7 (1943) 184–202.
43. ______, Isomorphisms of direct decompositions II, Bull. Acad. Sci. USSR, 10 (1946) 47–72.
44. ______, The Theory of Groups, translated by K. A. Hirsch from the second Russian edition, vol. 1, 2, Chelsea, New York, 1956.
45. L. Lovasz, On the cancellation law among finite relational structures, Periodic. Math. Hungar., 1 (1971) 145–156.
46. Magnus, Karrass, Solitar, Combinatorial Group Theory, Dover Publications, 1966.
47. R. McKenzie, A method for obtaining refinement theorems with an application to direct products of semigroups, Algebra Universalis, 2 (1972) 324–338.
48. G. S. Monk, A characterization of exchange rings, Proc. Amer. Math. Soc., 35 (1972) 349–353.
49. O. Ore, Direct decompositions, Duke Math. J., 2 (1936) 581–596.
50. M. O. Rabin, Recursive unsolvability of group theoretic problems, Ann. of Math., 67 (1958) 172–194.
51. R. Remak, Uber die Zerlegunigen der endlichen gruppen in direkte unzerlegbure factoren, J. Reine Angew. Math., 139 (1911) 293–308.
52. S. Shastri, Some cancellation theorems, J. Austral. Math. Soc., 30 (1980) 87–97.
53. W. V. Vasconcelos, On local and stable cancellation, An. Acad. Brasil. Ciênc. 37 (1965) 389–393.
54. E. A. Walker, Cancellation in direct sums of groups, Proc. Amer. Math. Soc., 7 (1956) 898–902.
55. R. B. Warfield, Jr., Genus and cancellation for groups with finite commutator subgroups, J. Pure Appl. Algebra, 6 (1975) 125–132.
56. ______, Cancellation and exchange theorems for finitely generated nilpotent groups.
57. ______, A Krull-Schmidt theorem for infinite sums of modules, Proc. Amer. Math. Soc., 22 (1969) 460–465.
58. ______, Exchange rings and decompositions of modules, Math. Ann., 199 (1972) 31–36.
59. J. H. M. Wedderburn, On the direct product in the theory of finite groups, Ann. of Math., 10 (1909) 173–176.
60. J. Wiegold, Growth of finite groups, J. Austral. Math. Soc., 27 (1974) 133–143.

CALCULATOR FUNCTION APPROXIMATION

CHARLES W. SCHELIN
Department of Mathematics, University of Wisconsin-La Crosse, La Crosse, WI 54601

1. Introduction. For about the past 10 years elementary function values have been available at the touch of a button on a hand-held calculator. The unified algorithm used in most hand calculators to approximate elementary functions is the subject of this paper. Surprisingly, this general algorithm which employs bit by bit approximation does not depend on results or

---

The author received his Ph.D. in 1971 from the University of Montana working in numerical analysis under Charles Bryan. Since that time he has taught at UWL, spending summers as a research fellow at McDonnell Douglas, EXXON, NASA-Ames, and Iowa State University. This paper was written while the author was an American Council on Education fellow in academic administration at the University of Utah.

techniques of the calculus. To begin, a few excerpts from the history of elementary function approximation are presented.

2. Tabular Function Values. Prior to our era of automatic computation, most function values used in applied problems were read from tables. The literature of the eighteenth and nineteenth centuries abounds with interpolation schemes devised to extend tabular data. Even the use of a slide rule for three significant figure estimates of elementary function values amounts to a table reading exercise.

Trigonometric tables, in the form of arc-chord data, have been traced to the time of Hipparchus (150 B.C.) through the work of Ptolemy (150 A.D.). Ptolemy's table, the result of meticulous measurements of inscribed regular polygons and an arc addition formula, was used by astronomers for better than one-thousand years [7]. The arc addition formula may be considered as the basis for today's trigonometric addition formulas.

Henry Briggs (1561–1631), credited with the base-first approach to logarithms, presented the properties as well as applications, a method of construction, and the first table of common logarithms in his Arithmetica Logarithmica (1624). This work and a later table containing sines, tangents, log sines, and log tangents, from 0° to 90°, in increments of one-hundredth of a degree, and with function values to fifteen decimal places, served as the standard for all later refinements. Of particular interest here is Briggs' computation scheme and an extension of that scheme suggested three hundred years later by Glaisher [4]. An outline of these procedures serves as a revealing introduction to the unified scheme used on today's hand calculators.

After defining the logarithms of 1 and 10 to be 0 and 1, respectively, Briggs began extracting successive square roots of 10, reasoning that each such extraction corresponded to halving the corresponding logarithm. Desiring 15 figure accuracy, he carried 32 decimal places and continued until the root extracted was within 10^{-15} of unity. This tedious process required 54 such extractions and yielded a skeletal table of values $\alpha_0, \alpha_1, \ldots, \alpha_{54}$, where $\alpha_0 = 10$, and

$$\alpha_{k+1} = \sqrt{\alpha_k}, \quad k = 0, 1, \ldots, 53, \tag{2.1}$$

thus

$$\log \alpha_k = 2^{-k}, \quad k = 0, 1, \ldots, 54.$$

To use these skeletal values, Briggs asserted that if $\alpha = 1 + a$, and $\beta = 1 + b$, with $0 < a < 10^{-15}$, and $0 < b < 10^{-15}$, then

$$\log \beta \doteq \frac{\log \alpha}{a} b$$

is accurate to 15 significant figures. Using the value of a from α_{54} (i.e., $10^{2^{-54}} - 1$), he obtained $(\log(1 + a))/a = .43429\ldots$ as the constant of proportionality. Now for any prime number P, square roots were extracted until $\beta = 1 + b$, with $b < 10^{-15}$, was obtained, then $\log \beta$ was taken as $(.43429\ldots)b$. Finally, $\log \beta$ was doubled the same number of times as square roots were extracted, the end result being a 15 figure approximation of $\log P$. Composite integers N were merely factored into primes whose logarithms were summed to obtain $\log N$.

Today, using results of the calculus that were developed more than a century after Briggs' work, one can readily establish his proportionality claim. Clearly, $(\log(1 + a))/a \to \log e$, as $a \to 0$, and so the constant of proportionality used in these calculations was an approximation to $\log e$. The accuracy claims can be easily substantiated as well. Using the Maclaurin series expansion for $\ln(1 + x)$, one can show that for $0 < a < 1$ and $0 < b < 1$,

$$|\log(1 + b) - \frac{b}{a}\log(1 + a)| < M^2/\ln 100$$

where $M = \max\{a, b\}$.

In 1916 Glaisher offered a method of resolving any positive real number into a product of

certain of the Briggs skeletal values given in (2.1), and perhaps powers of 10. By summing the corresponding logarithms of the α_j's used in this product, the desired logarithm is obtained without performing the root extractions or using the constant of proportionality. To resolve a number x, between 1 and 10, proceed by dividing x by the largest α_j less than x, then divide the quotient by the largest α_j less than itself, and so on until unity is reached (to 15 decimal places). When completed with this process, x may be written as a product of the α_j used as divisors, and thus $\log x$ is the sum of the corresponding $\log \alpha_j$. Glaisher concludes by noting that the required logarithm has base 2 representation

$$\sum_{k=0}^{54} \delta_k 2^{-k}, \text{ where } \delta_k = \begin{cases} 1, & \text{if } \alpha_k \text{ is used in the product representation,} \\ 0, & \text{otherwise.} \end{cases}$$

The idea of decomposing a given quantity into a combination (product here) of cleverly prescribed constants is fundamental to today's calculator schemes.

With the development of the calculus, series expansion became the primary method of approximating elementary functions. In 1669 Newton's De Analysis [**16**] contained series for the sine, cosine, and exponential functions. A year earlier, series for the natural logarithm and inverse tangent were published by Mercator and Gregory, respectively. By the end of the nineteenth century, through the work of such men as Fourier, Chebyshev, and Weierstrass, rigor had been added to function approximation theory.

Table refinements, using series expansions, continued through the first half of the twentieth century. Among the more recent efforts was the 1939 Project for the Computation of Mathematical Tables, sponsored by the National Bureau of Standards, and funded through the Works Progress Administration. (See, for example, [**10**].) After more than 300 years of fine tuning, economy of size was probably the primary improvement over the early function value tables. Domain reduction techniques, depending on function properties and notational advances, allowed a decrease in the number of tabular entries without decreasing the domain of their application. Today, domain reduction techniques are fundamental to the automatic computation of function values.

3. Computer Function Evaluation. Elementary function values were first generated electronically during the 1940's as inputs to analogue computations. In analogue computers numerical quantities are represented as voltages. By applying electrical devices, the equivalent of arithmetic operations are performed on the voltages. Among the analogue multiplication schemes is the servo multiplier. With this scheme a variable voltage is effectively multiplied by a constant by applying the voltage across a potentiometer, adjusted to the prescribed constant value. Using a resolver, a specially wound potentiometer, a voltage proportional to $\sin\theta$, or $\cos\theta$, is similarly obtained from a given voltage θ. Such generation is used to convert between rectangular and polar coordinates, and thus voltages representing inverse tangents are produced in a continuous fashion as well.

In digital computers numerical quantities are represented digit by digit, actually bit by bit, rather than by the intensity of an electromagnetic force. The computation circuitry for these machines is designed to perform only the four arithmetic operations. Function evaluation schemes, therefore, are dependent upon the efficient and accurate execution of an appropriate sequence of the four arithmetic operations. Typically, an evaluation subroutine employs a near minimax polynomial, or rational function, approximation on a selected interval. Given domain values are reduced to the interval of approximation, then stored coefficients are applied to the reduced domain value to produce the approximation. The scheme used to evaluate e^x in the IBM 360 Scientific Subroutine package illustrates this procedure. First the rational function

$$R(x) = 1 - \cfrac{2x}{.034657359x^2 + x + 9.9545957821 - \cfrac{617.9722695329}{x^2 + 87.4174972022}}$$

is used to approximate 2^{-x} on $[0, 1]$, with maximum relative error 2×10^{-9}. Now, to find e^x for a given x, take $y = -x \log_2 e$, ($\log_2 e = 1.442695041$ to 10 significant figures), $n = [y]$, ($[y]$ being the greatest integer less than or equal to y), and $z = y - n$. Then, $0 \leqslant z < 1$, and

$$e^x \doteq 2^{-n} R(z).$$

When working in base 2, 2^{-n} merely represents an n-bit shift of the binary point. Approximations used on various digital computers through the mid 1960's are contained in Fike [**2**], Hastings [**5**], and Kogbetliantz [**8**].

4. The CORDIC Scheme. By the early 1970's, miniaturization of electronics circuitry had advanced to a level which allowed the inclusion of function evaluation capabilities on hand calculators. A first difference between calculator circuitry and digital computer circuitry was the binary-coded decimal (BCD) representation of numbers. Hardware for conversion between the user's base 10 preference and the off-on switching simplicities of base 2 was not practical because of cost limitations and the fact that only a few arithmetic operations are used with each calculator input/output [**12**]. In BCD each digit of a base 10 numeral is represented in binary form, and the arithmetic operations, at least addition and subtraction, are wired accordingly. In choosing the function approximation schemes, calculation speed required that divisions and multiplications be avoided, and size dictated the use of a uniform scheme applicable to various elementary functions. For speed and efficiency, Hewlett-Packard selected the CORDIC (Coordinate Rotation Digital Computer) scheme for trigonometric function evaluation on the HP-35 [1]. Today, Hewlett-Packard, Texas Instruments, and many other manufacturers use some version of this scheme for all function evaluation.

The CORDIC computing technique was first employed by Volder in 1959, [**14**], to solve the trigonometric relationships that arise in navigation problems. Involving only a fixed sequence of additions, or subtractions, and binary shifts, this scheme was used to quickly and systematically approximate the value of a trigonometric function or its inverse. Following generalizations to base 10 arithmetic [**13**], and to other elementary functions [**15**], a CORDIC arithmetic processor chip is now available to perform multiplications and divisions, to calculate square roots, and to evaluate the sine, cosine, tangent, arctangent, sinh, cosh, tanh, arctanh, ln, and exp functions [**6**]. In binary form the unified CORDIC scheme consists only of the iterative equations

$$x_{k+1} = x_k - m\,\delta_k y_k 2^{-k}, \tag{4.1}$$

$$y_{k+1} = y_k + \delta_k x_k 2^{-k}, \tag{4.2}$$

$$z_{k+1} = z_k - \delta_k \varepsilon_k, \tag{4.3}$$

$$\delta_k = \pm 1, \quad \text{for } k = 0, 1, \ldots, n, \tag{4.4}$$

where $m = 1$, 0, or -1, is a mode indicator and $\{\varepsilon_k\}$ is a sequence of prestored constants depending on m. By appropriately selecting x_0, y_0, z_0, and the sign of each δ_k, these four equations may be used to generate approximations for any of the elementary functions mentioned above.

Before deriving the CORDIC scheme, consider the following decomposition theorem which summarizes observations of Volder [**14**] and Walther [**15**], and upon which the scheme depends.

THEOREM. *Suppose $\varepsilon_0 \geqslant \varepsilon_1 \geqslant \varepsilon_2 \geqslant \cdots \geqslant \varepsilon_n > 0$ is a finite sequence of real numbers such that*

$$\varepsilon_k \leqslant \sum_{j=k+1}^{n} \varepsilon_j + \varepsilon_n, \quad \text{for } 0 \leqslant k \leqslant n, \tag{i}$$

and suppose r is a real number such that

$$|r| \leqslant \sum_{j=0}^{n} \varepsilon_j. \tag{ii}$$

If $s_0 = 0$, *and* $s_{k+1} = s_k + \delta_k \varepsilon_k$, *for* $0 \leqslant k \leqslant n$, *where*

(iii) $$\delta_k = \begin{cases} 1, & \text{if } r \geqslant s_k, \\ -1, & \text{if } r < s_k, \end{cases}$$

then

(iv) $$|r - s_k| \leqslant \sum_{j=k}^{n} \varepsilon_j + \varepsilon_n, \quad \text{for } 0 \leqslant k \leqslant n,$$

and so in particular $|r - s_{n+1}| < \varepsilon_n$.

Proof. (Induction on k). For $k = 0$,

$$|r - s_0| = |r| \leqslant \sum_{j=0}^{n} \varepsilon_j < \sum_{j=0}^{n} \varepsilon_j + \varepsilon_n,$$

using (ii). Assuming $|r - s_k| \leqslant \Sigma_{j=k+1}^{n} \varepsilon_j + \varepsilon_n$, consider $|r - s_{k+1}|$. Note that from (iii), δ_k and $r - s_k$ have the same sign and thus

$$|r - s_{k+1}| = |r - s_k - \delta_k \varepsilon_k| = ||r - s_k| - \varepsilon_k|.$$

Now,

$$|r - s_k| - \varepsilon_k \leqslant \sum_{j=k}^{n} \varepsilon_j + \varepsilon_n - \varepsilon_k = \sum_{j=k+1}^{n} \varepsilon_j + \varepsilon_n,$$

by the inductive hypothesis. Also, from (i),

$$-\left[\sum_{j=k+1}^{n} \varepsilon_j + \varepsilon_n\right] \leqslant -\varepsilon_k,$$

and so

$$-\left[\sum_{j=k+1}^{n} \varepsilon_j + \varepsilon_n\right] \leqslant |r - s_k| - \varepsilon_k.$$

Combining these two results,

$$||r - s_k| - \varepsilon_k| \leqslant \sum_{j=k+1}^{n} \varepsilon_j + \varepsilon_n,$$

and the theorem is true.

Note that each approximant ε_j is used in the decomposition described above. This approach is favored over the seemingly more natural option of allowing $\delta_j = 0$ in some cases, because of the associated circuit simplicity. Condition (i) allows the use of all ε_j. Should $|r - s_k|$ be less than ε_n for some $k < n$, then ε_k can be added and the remaining ε_j subtracted, returning again to $|r - s_{n+1}| < \varepsilon_n$.

To obtain the CORDIC scheme, suppose θ is decomposed in the fashion described above, and the trigonometric addition formulas are applied to the approximation $s_{k+1} = s_k + \delta_k \varepsilon_k$. Then

(4.5) $$\begin{aligned} \cos(s_{k+1}) &= \cos s_k \cos \varepsilon_k - \delta_k \sin s_k \sin \varepsilon_k \\ &= \cos \varepsilon_k (\cos s_k - \delta_k \sin s_k \tan \varepsilon_k) \end{aligned}$$

and

(4.6) $$\begin{aligned} \sin(s_{k+1}) &= \sin s_k \cos \varepsilon_k + \delta_k \cos s_k \sin \varepsilon_k \\ &= \cos \varepsilon_k (\sin s_k + \delta_k \cos s_k \tan \varepsilon_k). \end{aligned}$$

Now, if $s_0 = 0$, and $\varepsilon_k = \tan^{-1} 2^{-k}$ for $0 \leqslant k \leqslant n$, and if x_{k+1} and y_{k+1} are associated with

$\cos(s_{k+1})$ and $\sin(s_{k+1})$, respectively, then but for the $\cos \varepsilon_k$ factor, equations (4.5) and (4.6) could be replaced by (4.1) and (4.2) with $m = 1$. Actually, if $x_0 = 1, y_0 = 0$, and $K = \Pi_{j=0}^{n} \cos \varepsilon_j$, then $\cos(s_{n+1}) = Kx_{n+1}$, and $\sin(s_{n+1}) = Ky_{n+1}$. So by choosing $x_0 = 1/K$ and $y_0 = 0$, we obtain $x_{n+1} = \cos(s_{n+1})$, and $y_{n+1} = \sin(s_{n+1})$ from (4.1) and (4.2), with $m = 1$.

To verify that the choice $\varepsilon_k = \tan^{-1} 2^{-k}$ satisfies hypothesis (i) of the decomposition theorem, the mean value theorem may be applied to the arctangent function to obtain

$$\varepsilon_j - \varepsilon_{j+1} \leqslant 2^{j+1}/(1 + 2^{2j+2})$$

and

$$\varepsilon_j \geqslant 2^j/(1 + 2^{2j}), \quad \text{for } j = 0, 1, \ldots, n.$$

Thus

$$\varepsilon_k - \varepsilon_n = \sum_{j=k}^{n-1} (\varepsilon_j - \varepsilon_{j+1}) \leqslant \sum_{j=k+1}^{n} 2^j/(1 + 2^{2j}) \leqslant \sum_{j=k+1}^{n} \varepsilon_j,$$

establishing the desired result.

Clearly $\Sigma_{j=0}^{3} \varepsilon_j > \pi/2$, and so from (ii) of the decomposition theorem we need only require that $|\theta| \leqslant \pi/2$. Domain reduction to this interval may be achieved through division of a given θ by $\pi/2$, or through repeated subtractions of π (when $\theta > 0$).

In addition to equations (4.1) and (4.2), the complete CORDIC scheme includes the decomposition equation (4.3), in which the approximation s_{n+1} to a given θ is formed, as well as the assignment equation (4.4). Taking $z_0 = \theta$, and

$$\delta_k = \begin{cases} 1, & \text{if } z_k \geqslant 0 \\ -1, & \text{if } z_k < 0, \end{cases}$$

the s_k of the decomposition theorem are related to the z_k of (4.3) by $z_k = \theta - s_k$. Here the δ_k are selected to force z_k toward zero, hence forcing $\Sigma_{j=0}^{n} \delta_k \varepsilon_k$ toward θ. Volder termed this selection of the δ_k the *rotation* mode of the CORDIC scheme. This title is descriptive for in the rotation mode we begin with (x_0, y_0) on the x-axis and rotate $\pm\varepsilon_k$ radians in the kth step until after n steps, (x_{n+1}, y_{n+1}) lies on the terminal side of an angle whose radian measure is within ε_n of the given θ. (See Fig. 1.)

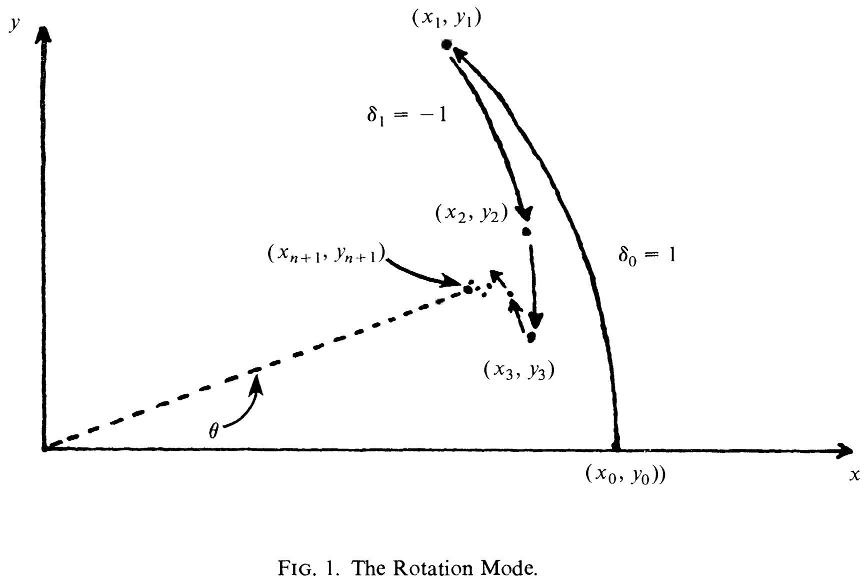


FIG. 1. The Rotation Mode.

A first indication of the versatility of the CORDIC scheme is encountered in the computation of inverse tangents. To find $\theta = \tan^{-1} w$ for a given w, we think of w as y_0/x_0 where (x_0, y_0) is on the terminal side of θ. Then the δ_k are selected so that the terminal side of θ is rotated to the x-axis. The negative of these accumulated rotations forms θ, within ε_n, and is recorded as z_{n+1} if we begin with $z_0 = 0$. Volder terms this selection of δ_k to force y_k toward zero, i.e.,

$$\delta_k = \begin{cases} 1, & \text{if } y_k < 0 \\ -1, & \text{if } y_k \geqslant 0, \end{cases}$$

the *vectoring* mode of the CORDIC scheme. (See Figure 2.) In the vectoring mode, if w is given, $z_0 = 0, y_0$ is selected to satisfy (ii) of the decomposition theorem, and $x_0 = y_0/w$, then

$$z_{n+1} \doteq \tan^{-1} w \quad \text{and} \quad x_{n+1} \doteq K\sqrt{x_0^2 + y_0^2}\,.$$

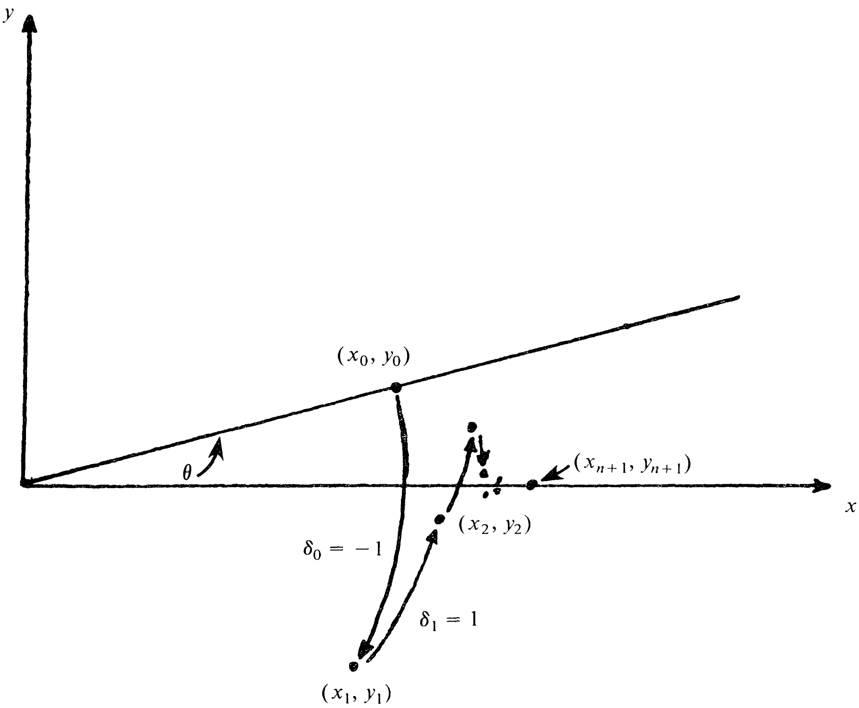


FIG. 2. The Vectoring Mode.

Recall that in today's hand calculators, base 10 arithmetic is employed, so in actuality $\varepsilon_k = \tan^{-1} 10^{-k}$, and equations (4.1) through (4.4) are evaluated 10 times for each k.

In moving to hyperbolic functions, the application of hyperbolic addition formulas to approximation s_{k+1} would leave (4.2) unchanged and require that $m = -1$ in equation (4.1), if x_j and y_j are associated with $\cosh \varepsilon_j$ and $\sinh \varepsilon_j$, respectively, and if $\varepsilon_k = \tanh^{-1} 2^{-k}$. Unfortunately, in this case the ε_k do not satisfy (i) of the decomposition theorem for all k, and so for certain k the equations (4.1) through (4.4) must be evaluated twice [**15**]. With this in mind, if $K' = \Pi \cosh \varepsilon_j$, where the product contains the needed duplications, $x_0 = 1/K'$, $y_0 = 0$, $z_0 = \theta$, and the δ_k are selected for the rotation mode, then $x_{n+1} \doteq \cosh\theta$, and $y_{n+1} \doteq \sinh\theta$, so

$$e^{\theta} \doteq x_{n+1} + y_{n+1}.$$

In the hyperbolic vectoring mode, if x_0 and y_0 are given and $z_0 = 0$, then

$$z_{n+1} \doteq \tanh^{-1}(y_0/x_0) \quad \text{and} \quad x_{n+1} \doteq K'\sqrt{x_0^2 + y_0^2}\,.$$

Perhaps the most useful applications of this mode are realized when for a given w, $x_0 = w + 1$ and $y_0 = w - 1$, or $x_0 = w + .25$ and $y_0 = w - .25$. In the first case we obtain

$$\ln w \doteq 2z_{n+1},$$

while the latter case yields

$$\sqrt{w} \doteq x_{n+1}/K'.$$

A final surprising feature of the CORDIC scheme is its utility in performing multiplications and divisions if equation (4.1) is altered by setting $m = 0$, and $\varepsilon_k = 2^{-k}$. In the rotation mode, if $y_0 = 0$, and x_0 and z_0 are given, $\Sigma_{k=0}^{n}\delta_k\varepsilon_k$ approximates z_0 thus

$$y_{n+1} \doteq x_0 z_0.$$

In the vectoring mode, if $z_0 = 0$, and x_0 and y_0 are given, then $y_0 + x_0[\Sigma_{k=0}^{n}\delta_k\varepsilon_k]$ is forced to zero, hence

$$z_{n+1} = -\sum_{k=0}^{n} \delta_k\varepsilon_k \doteq y_0/x_0.$$

The multiplication-division form of the CORDIC scheme is analogous to the pseudo multiplication and division of Meggitt [**11**]. A comparison of the techniques of Meggitt, Volder, and others, with error analyses, can be found in Franke [**3**]. Numerical examples of CORDIC computations appear in Kropa [**9**].

The various modes of the generalized CORDIC scheme are summarized in the table below.

The Generalized Binary CORDIC Scheme

$$\begin{aligned} x_{k+1} &= x_k - m\delta_k y_k 2^{-k} \\ y_{k+1} &= y_k + \delta_k x_k 2^{-k} \qquad \text{for } k = 0, 1, \ldots, n. \\ z_{k+1} &= z_k - \delta_k\varepsilon_k \end{aligned}$$

	Rotation ($z_k \to 0$) $\delta_k = \begin{cases} 1, & \text{if } z_k \geqslant 0 \\ -1, & \text{if } z_k < 0 \end{cases}$	Vectoring ($y_k \to 0$) $\delta_k = \begin{cases} 1, & \text{if } y_k < 0 \\ -1, & \text{if } y_k \geqslant 0 \end{cases}$
$m = 0$ $\varepsilon_k = 2^{-k}$	x_0 given, $y_0 = 0$, z_0 given yields $y_{n+1} \doteq x_0 z_0$	x_0 given, y_0 given, $z_0 = 0$ yields $z_{n+1} \doteq y_0/x_0$
$m = 1$ $\varepsilon_k = \tan^{-1}2^{-k}$ $K = \Pi \cos\varepsilon_k$	$x_0 = 1/K, y_0 = 0, z_0 = \theta$ yields $x_{n+1} \doteq \cos\theta, y_{n+1} \doteq \sin\theta$	x_0 given, y_0 given, $z_0 = 0$ yields $z_{n+1} \doteq \tan^{-1} y_0/x_0$, and $x_{n+1} \doteq K\sqrt{x_0^2 + y_0^2}$
$m = -1$ $\varepsilon_k = \tanh^{-1}2^{-k}$ $K' = \Pi \cosh\varepsilon_k$	$x_0 = 1/K', y_0 = 0, z_0 = \theta$ yields $x_{n+1} \doteq \cosh\theta, y_{n+1} \doteq \sinh\theta$, $e^\theta \doteq x_{n+1} + y_{n+1}$	x_0 given, y_0 given, $z_0 = 0$ yields $z_{n+1} \doteq \tanh^{-1} y_0/x_0$, and $x_{n+1} \doteq K'\sqrt{x_0^2 - y_0^2}$ $x_0 = w + 1, y_0 = w - 1$ yields $\ln w \doteq 2z_{n+1}$ $x_0 = w + .25, y_0 = w - .25$ yields $\sqrt{w} \doteq x_{n+1}/K'$

References

1. D. Cochran, Algorithms and accuracy in the HP-35, Hewlett-Packard J. (June 1972) 10–11.
2. C. Fike, Computer Evaluation of Mathematical Functions, Prentice-Hall, Englewood Cliffs, NJ, 1968.
3. R. Franke, An Analysis of Algorithms for Hardware Evaluation of Elementary Functions, National Technical Information Service Publication no. AD 761519, U.S. Department of Commerce, 1973.
4. J. Glaisher, On some methods of calculating logarithms without the use of series, Quart. J. Pure and Appl. Math., v. 17, (1917) 249–301.
5. C. Hastings, Approximations for Digital Computers, Princeton University Press, Princeton, NJ, 1955.
6. G. Haviland and A. Tuszynski, A CORDIC arithmetic processor chip, IEEE Trans. Computers, v. C-29, no. 2 (February 1980) 68–79.
7. M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York, 1972, p. 125.
8. E. Kogbetliantz, Generation of Elementary Functions, in Mathematical Methods for Digital Computers, A. Ralston and H. Wilf, Editors, Wiley, New York, 1960.
9. J. Kropa, Calculator algorithms, Math. Mag., v. 51, no. 2 (March 1978) 106–109.
10. A. Lowan, Editor, Tables of the Exponential Function e^x, Government Printing Office, Washington, D.C., 1939.
11. J. Meggit, Pseudo division and pseudo multiplication processes, IBM J. (April 1962) pp. 210–226.
12. H. Schmid, Decimal Computation, Wiley, New York, 1974, p. 10.
13. H. Schmid and A. Bogocki, Use decimal CORDIC for generation of many transcendental functions, Electrical Design News Magazine (February 1973) 64–73.
14. J. Volder, The CORDIC computing technique, IRE Trans. Computers, v. EC-8 (September 1959) 330–334.
15. J. Walther, A Unified Algorithm for Elementary Functions, Joint Computer Conference Proceedings, v. 38, Spring 1971, pp. 379–385.
16. D. Whiteside, Editor, The Mathematical Works of Isaac Newton, v. 1, Johnson Reprint Corp., New York, 1964.

HOW SMALL CAN THE MEAN SHADOW OF A SET BE?

ALLEN J. SCHWENK

Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402

J. IAN MUNRO

Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Consider the following problem concerning projections of finite sets:

Given n points in Euclidean d-space and the projections onto the $\binom{d}{s}$ hyperplanes defined by subsets of s of the coordinate axes, what is the minimum number of distinct points in the projection of greatest cardinality?

As a student at Caltech, Allen Schwenk won the E.T. Bell Undergraduate Research Prize. He received his Ph.D. from the University of Michigan in 1973. He is now an associate professor at the U.S. Naval Academy where he has spent most of his career, except for single years as a NATO Postdoctoral Fellow, University of Oxford (1973–74), at Michigan State (1974–75), and recently in the Department of Combinatorics and Optimization, University of Waterloo (1981–82). His research interest is primarily in graph theory, specifically the spectrum of a graph and graphical enumeration, with some occasional spill over into combinatorics. His hobbies include duplicate bridge, bicycling, Rubik's cube, and jogging. Beat Army.

Ian Munro is a Professor in the Computer Science Department at the University of Waterloo where he has been since 1971. His major research interest is the efficiency of data structures. Although this paper is indicative of his interest in combinatorics, it was a problem in data structuring which gave rise to the result discussed here.

References

1. D. Cochran, Algorithms and accuracy in the HP-35, Hewlett-Packard J. (June 1972) 10–11.
2. C. Fike, Computer Evaluation of Mathematical Functions, Prentice-Hall, Englewood Cliffs, NJ, 1968.
3. R. Franke, An Analysis of Algorithms for Hardware Evaluation of Elementary Functions, National Technical Information Service Publication no. AD 761519, U.S. Department of Commerce, 1973.
4. J. Glaisher, On some methods of calculating logarithms without the use of series, Quart. J. Pure and Appl. Math., v. 17, (1917) 249–301.
5. C. Hastings, Approximations for Digital Computers, Princeton University Press, Princeton, NJ, 1955.
6. G. Haviland and A. Tuszynski, A CORDIC arithmetic processor chip, IEEE Trans. Computers, v. C-29, no. 2 (February 1980) 68–79.
7. M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York, 1972, p. 125.
8. E. Kogbetliantz, Generation of Elementary Functions, in Mathematical Methods for Digital Computers, A. Ralston and H. Wilf, Editors, Wiley, New York, 1960.
9. J. Kropa, Calculator algorithms, Math. Mag., v. 51, no. 2 (March 1978) 106–109.
10. A. Lowan, Editor, Tables of the Exponential Function e^x, Government Printing Office, Washington, D.C., 1939.
11. J. Meggit, Pseudo division and pseudo multiplication processes, IBM J. (April 1962) pp. 210–226.
12. H. Schmid, Decimal Computation, Wiley, New York, 1974, p. 10.
13. H. Schmid and A. Bogocki, Use decimal CORDIC for generation of many transcendental functions, Electrical Design News Magazine (February 1973) 64–73.
14. J. Volder, The CORDIC computing technique, IRE Trans. Computers, v. EC-8 (September 1959) 330–334.
15. J. Walther, A Unified Algorithm for Elementary Functions, Joint Computer Conference Proceedings, v. 38, Spring 1971, pp. 379–385.
16. D. Whiteside, Editor, The Mathematical Works of Isaac Newton, v. 1, Johnson Reprint Corp., New York, 1964.

HOW SMALL CAN THE MEAN SHADOW OF A SET BE?

ALLEN J. SCHWENK
Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402

J. IAN MUNRO
Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Consider the following problem concerning projections of finite sets:

Given n points in Euclidean d-space and the projections onto the $\binom{d}{s}$ hyperplanes defined by subsets of s of the coordinate axes, what is the minimum number of distinct points in the projection of greatest cardinality?

As a student at Caltech, Allen Schwenk won the E.T. Bell Undergraduate Research Prize. He received his Ph.D. from the University of Michigan in 1973. He is now an associate professor at the U.S. Naval Academy where he has spent most of his career, except for single years as a NATO Postdoctoral Fellow, University of Oxford (1973–74), at Michigan State (1974–75), and recently in the Department of Combinatorics and Optimization, University of Waterloo (1981–82). His research interest is primarily in graph theory, specifically the spectrum of a graph and graphical enumeration, with some occasional spill over into combinatorics. His hobbies include duplicate bridge, bicycling, Rubik's cube, and jogging. Beat Army.

Ian Munro is a Professor in the Computer Science Department at the University of Waterloo where he has been since 1971. His major research interest is the efficiency of data structures. Although this paper is indicative of his interest in combinatorics, it was a problem in data structuring which gave rise to the result discussed here.

We will demonstrate that not only this min-max value, but indeed the geometric mean of these $\binom{d}{s}$ cardinalities is at least $n^{s/d}$. This lower bound on the geometric mean is achieved when the points are arranged in any hypercuboid of lattice points orthogonal to the axes (including the trivial pattern of a row of n points along one axis). The min-max is certainly achievable if n happens to be a perfect dth power. The complementary problem of determining an upper bound for mean projection size is trivial as the points $\{(i, i, \ldots, i): 1 \leqslant i \leqslant n\}$ do not compress under projections.

This problem arose [1] in establishing a lower bound on the number of comparisons necessary to find a d-key record which has been stored in an array of n such records and is specified by s of its keys. We feel the problem is of interest in its own right as a fundamental result about projections of a finite set, and are pleased that it can be demonstrated by elementary techniques requiring only multiple induction and a finite form of Hölder's Inequality.

We now introduce the required notation and formally state the theorem.

Let $X = \{x_1, \ldots, x_n\}$ denote an arbitrary (finite) set of n distinct points in Euclidean d-space, E^d, and let $N_d = \{1, \ldots, d\}$ be the index set representing the coordinates. We wish to consider projections onto a subspace E^s with $s < d$ effected by retaining s coordinates from each point and deleting the other $d - s$ coordinates. There are $\binom{d}{s}$ possible projections for us to choose from. We identify a particular projection by selecting $S \subset N_d$ with $|S| = s$. For example, $S = N_s \subset N_d$ identifies the projection that keeps the first s coordinates and discards the last $n - s$. Of course discarding some coordinates might leave some of the x_i's identical, that is, different points can project to the same image point in E^s. We let $\pi(X, S)$ denote the set of image points that result when the set X is projected onto the subspace E^s whose s coordinates are given by S. Let $p(X, S)$ be the order of $\pi(X, S)$; clearly $1 \leqslant p(X, S) \leqslant n$. We shall think of these projections as "shadows" of the original set. This term is most appropriate when $s = d - 1$. For other values of s the term may be a bit contrived, but it still provides a perspective which helps one recall the result.

The original problem is to find the minimum over all n-sets X of the maximum over all s-sets S of $p(X, S)$. Our theorem is somewhat stronger in that it provides a sharp lower bound on the geometric mean of all the projections.

THEOREM. *For any set X of n points in E^d, the projection size is bounded by*

$$\text{(1)} \qquad \text{Geom. Mean}\,\{p(X, S): S \subset N_d \text{ and } |S| = s\} \geqslant n^{s/d}.$$

Furthermore, if n can be factored as $n = a_1 a_2 \ldots a_d$ with each a_i a natural number, then the bound is attained by choosing X to be the Cartesian product $\bigtimes_{i=1}^{d} N_{a_i}$.

Proof. We first verify that the specified Cartesian product attains the bound. Evidently, $p(X, S) = \prod_{i \in S} a_i$. Since each i lies in $\binom{d-1}{s-1}$ different s-sets, we have

$$\text{(2)} \qquad \prod_{|S|=s} p(X, S) = \prod_{i=1}^{d} a_i^{\binom{d-1}{s-1}} = n^{\binom{d-1}{s-1}}.$$

Taking the $\binom{d}{s}$th root of this equation yields

$$\text{(3)} \qquad \text{Geom. Mean}\,\{p(X, S): S \subset N_d \text{ and } |S| = s\} = n^{s/d}.$$

Notice that a suitable factorization exists for all n and d, since $a_1 = n$ and $a_i = 1$ for $i \geqslant 2$ provides a trivial factorization in every case.

To prove the bound, we observe that the geometric mean involves a $\binom{d}{s}$th root, so we raise both sides to the $\binom{d}{s}$ power to get

$$\text{(4)} \qquad \prod_{|S|=s} p(X, S) \geqslant n^{\binom{d-1}{s-1}}.$$

We must verify this inequality to prove the bound. First, we shall treat the case $s = d - 1$ with the proof obtained by induction on d and depending on the lemma presented below. Then we complete the proof for arbitrary s by induction on the difference $d - s$. First, the lemma:

LEMMA (*Hölder's Inequality*). *For all positive reals x_{ij} and positive integer p,*

$$\prod_{j=1}^{p} \sum_{i=1}^{k} x_{ij} \geqslant \left(\sum_{i=1}^{k} \prod_{j=1}^{p} x_{ij}^{1/p} \right)^{p}.$$

Proof. This is in essence a finite variation of Hölder's Inequality reported as inequality (2.7.2) in Hardy, Littlewood, and Pólya [**2**, page 22]. For our purposes, we have specialized (2.7.2) by selecting $\alpha = \beta = \cdots = \lambda = 1/p$ and we have raised both sides to the pth power. The more general form of (2.7.2) requires only that $\alpha, \beta, \ldots, \lambda$ be p positive variables whose sum is 1. The reader might notice that the case $p = 1$ gives a trivial equality while $p = 2$ is just a version of the familiar Cauchy-Swarz inequality. The inequalities for $p \geqslant 3$ are less well known.

We now return to the proof of the theorem for the case $s = d - 1$, and we proceed by induction on d. The first case, $d = 2$, is essentially solved by the pigeonhole principle. For $i = 1$ and 2, let $S_i = N_d - \{i\}$ be the set that omits i. (This strange choice of notation is made for later convenience.) Clearly, every point in X lies in the set $\pi(X, S_2) \times \pi(X, S_1)$. Thus $p(X, S_1)p(X, S_2) \geqslant n$ as required by (4). Now assume the inequality has been verified for dimension $d - 1$ and consider the d-dimensional problem. Denote $p(X, S_1)$ by r. Thus the n points of X lie along r lines parallel to the 1st coordinate axis. The n points also determine k distinct hyperplanes of dimension $d - 1$ which are orthogonal to this 1st axis. Since the n points lie among the rk points found by intersecting lines with hyperplanes, we have

$$rk \geqslant n. \tag{5}$$

Let X_i denote the set of points in the ith hyperplane and set $n_i = |X_i|$. Clearly $1 \leqslant n_i \leqslant r$ and $\sum_{i=1}^{k} n_i = n$. In each hyperplane, there are $d - 1$ different projections onto $(d - 2)$-dimensional subspaces. Let T_j for $2 \leqslant j \leqslant d$ denote the index set $T_j = N_d - \{1, j\}$. Now $p(X_i, T_j)$ counts the order of the image of the projection from the ith hyperplane onto the coordinates of T_j. But now we replace the first coordinate and consider the corresponding projection onto S_j. Because the hyperplanes form a parallel family, no image point can have preimages in more than one hyperplane. Consequently, for $2 \leqslant j \leqslant d$,

$$p(X, S_j) = \sum_{i=1}^{k} p(X_i, T_j). \tag{6}$$

Recalling that $p(X, S_1) = r$, we are prepared to bound $\prod_{j=1}^{d} p(X, S_j)$. First, we substitute from (6) to get

$$\prod_{j=1}^{d} p(X, S_j) = r \prod_{j=2}^{d} \sum_{i=1}^{k} p(X_i, T_j). \tag{7}$$

Next we apply the Hölder Inequality with exponent $p = d - 1$ to find

$$\prod_{j=1}^{d} p(X, S_j) \geqslant r \left(\sum_{i=1}^{k} \prod_{j=2}^{d} \left(p(X_i, T_j) \right)^{1/(d-1)} \right)^{d-1}. \tag{8}$$

But by the induction hypothesis,

$$\prod_{j=2}^{d} p(X_i, T_j) \geqslant n_i^{d-2} \tag{9}$$

and so

$$\prod_{j=1}^{d} p(X, S_j) \geqslant r \left(\sum_{i=1}^{k} n_i^{(d-2)/(d-1)} \right)^{d-1}. \tag{10}$$

To simplify the lower bound, we begin by writing each term as $n_i \cdot n_i^{-1/(d-1)}$, so

$$(11)\qquad \prod_{j=1}^{d} p(X, S_j) \geqslant r\left(\sum_{i=1}^{k} n_i \cdot n_i^{-1/(d-1)}\right)^{d-1}.$$

Next, we distribute the leading factor of r into the sum. Because of the power to which the sum is raised, the factor r appears as $r^{1/(d-1)}$ in each term:

$$(12)\qquad \prod_{j=1}^{d} p(X, S_j) \geqslant \left(\sum_{i=1}^{k} n_i\left(\frac{r}{n_i}\right)^{1/(d-1)}\right)^{d-1}.$$

But we have already observed that $n_i \leqslant r$, so the quantity (r/n_i) is bounded below by 1. Thus, we may suppress this factor from each term to obtain

$$(13)\qquad \prod_{j=1}^{d} p(X, S_j) \geqslant \left(\sum_{i=1}^{k} n_i\right)^{d-1} = n^{d-1}.$$

Thus, (4) has been verified when $s = d - 1$ for all d.

The final stage of the proof is to induct on the difference $d - s$. Each set of size s is contained in $d - s$ different sets S_j of order $d - 1$. Whenever $S \subset S_j \subset N_d$ the projections $\pi(X, S)$ and $\pi(\pi(X, S_j), S)$ are identical sets. (The projection is accomplished in two stages in the second expression.) Consequently,

$$(14)\qquad \left[\prod_{|S|=s} p(X, S)\right]^{d-s} = \prod_{j=1}^{d} \prod_{\substack{|S|=s \\ S\subset S_j}} p(\pi(X, S_j), S).$$

But the inner product on the right represents a projection from E^{d-1} to E^s. By the induction hypothesis, we may insert the bound from inequality (4) to get

$$(15)\qquad \left[\prod_{|S|=s} p(X, S)\right]^{d-s} \geqslant \prod_{j=1}^{d} p(X, S_j)^{\binom{d-2}{s-1}} = \left[\prod_{j=1}^{d} p(X, S_j)\right]^{\binom{d-2}{s-1}}.$$

Finally, the theorem has already been verified for projections from E^d to E^{d-1} so we may bound the bracketted product on the right to conclude

$$(16)\qquad \left[\prod_{|S|=s} p(X, S)\right]^{d-s} \geqslant [n^{d-1}]^{\binom{d-2}{s-1}}.$$

Taking $d - s$ roots of both sides yields (4) completing the proof for arbitrary d and s.

Recall that the question which inspired this theorem sought a lower bound on max $\{p(X, S)$: $|S| = s$ and $S \subset N_d\}$. Since the maximum is an integer at least as large as the geometric mean, we may conclude $\max\{p(X, S): |S| = s \text{ and } S \subset N_d\} \geqslant \lceil n^{s/d}\rceil$. For which values of n, s, and d can we actually find a set X attaining the bound? When $s = 1$, set $m = \lceil n^{1/d}\rceil$. Let X be any subset of $\bigtimes_{i=1}^{d} N_m$ with order n. Clearly the maximum projection attains the bound of m. For $s > 1$, the conditions for attaining the bound become much more subtle. We illustrate with the smallest case, namely $s = 2$ and $d = 3$. By detailed and boring combinatorial argument, we find that $\lceil n^{2/3}\rceil$ is attained for all $n \leqslant 12$ except $n = 5, 10$, and 11 when the maximum for every X is at least $1 + \lceil n^{2/3}\rceil$ points. Certainly the best possible bound for other values of n, s, and d is a difficult combinatorial problem which we leave open for the highly ambitious. One final and trivial observation: Whenever $n = k^d$ happens to be a perfect dth power, the set $X = \bigtimes_{i=1}^{d} N_k$ achieves the bound for all s.

Throughout this discussion, we have limited our attention to projections onto coordinate axes only. However, it should be clear that any set of d mutually orthogonal directions could serve as the "coordinates" used in the theorem.

In conclusion, there is a continuous version of the theorem. Presumably, it could be proved directly from the integral form of Hölder's Inequality stated as inequality 188 in Hardy, Littlewood, and Pólya [**2**, p. 140] namely

$$\left(\prod_{j=1}^{p} \int f_j(x)\,dx\right)^{1/p} \geqslant \int \left(\prod_{j=1}^{p} f_j(x)\right)^{1/p} dx.$$

We state the continuous version as a corollary to the finite version using $m_d(X)$ to denote the d-dimensional measure of the set X.

COROLLARY. *If X is a set with finite positive Riemann measure,*

$$\text{Geom. Mean}\,\{m_s(\pi(X,S)) \colon |S| = s \text{ and } S \subset N_d\} \geqslant m_d(X)^{s/d}.$$

Sketch of proof. We shall replace the infinite set X by a sequence of finite sets X_k for each natural number k. Partition E^d into d-dimensional cubes with edge $1/k$ and measure $1/k^d$. Select the central point of each cube to lie in X_k if and only if the entire cell is contained in X. Now $m_d(X) \geqslant |X_k|/k^d$ and analogously $m_s(\pi(X,S)) \geqslant p(X_k,S)/k^s$. Take the geometric mean over all $S \subset N_d$ of order s and apply the theorem to the projections of X_k to obtain

$$\text{Geom. Mean}\,\{m_s(\pi(X,S)) \colon |S| = s \text{ and } S \subset N_d\} \geqslant |X_k|^{s/d}/k^s. \tag{17}$$

Since $\lim_{k\to\infty}|X_k|/k^d = m_d(X)$, the limit as $k \to \infty$ of the right-hand side is $m_d(X)^{s/d}$ as required.

Acknowledgment. The first author is grateful to the University of Waterloo and particularly to the Department of Combinatorics and Optimization for the hospitality extended to him during his recent sabbatical from the U.S. Naval Academy. The work of the second author was supported by NSERC grant A-8237.

References

1. H. Alt, K. Mehlhorn, and J. I. Munro, Partial match retrieval in implicit data structures, unpublished manuscript.

2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1967.

MISCELLANEA

103. Among the miscellaneous reading of my youth was a history of modern Europe, which concluded with a general survey and attempted forecast of progress in arts, science, and literature. So far as I can judge, this work was written about the time of Euler or Lagrange. On the subject of mathematics the writer's conclusion was that fruitful investigation seemed at an end, and that there was little prospect of brilliant discoveries in the future. To us, a century later, this judgment might seem to illustrate the danger of prophesying, and lead us to look upon the author as one who must have been too prone to hasty conclusions. I am not sure that careful analysis would not show the author's view to be less rash than it may now appear.

—SIMON NEWCOMB, address to the New York Mathematical Society, Bull. N.Y. Math. Soc., 3(1894) 95–107.

C E N T E R   S E C T I O N
(Vol. 90, No. 5, May 1983)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P. Actualités mathématiques. Jean-Paul Pier. Gauthier-Villars, 1982, 542 pp, 250 FF (P). [ISBN: 2-04-015427-2] Proceedings of the sixth congress of the organization of mathematicians speaking romance languages. JD-B

General, T(15-17), S*, L*. A Problem Seminar. Donald J. Newman. Springer-Verlag, 1982, vii + 113 pp, $12.95 (P). [ISBN: 0-387-90765-3] Over 100 intriguing problems at the advanced undergraduate- and graduate-level. Divided into three parts: problems, hints, solution sketches. The presentations are delightful: fresh, witty, informal, insightful. LCL

Mathematics Appreciation, S(13-15). L. Repunits and Repetends. Samuel Yates (157 Capri-D, Kings Point, Delray Beach, FL 33445), 1982, 215 pp, (P). The study of repunits and repetends has attracted both amateurs and professionals. This treatise, written chiefly for the former group, provides an introduction to this literature and puts forward a number of unsolved problems which might be explored on the microcomputer. LCL

Precalculus, T(13: 1).** Functions and Graphs: Calculus Preparatory Mathematics. Malcolm W. Pownall. Prentice-Hall, 1983, xv + 479 pp, $24.95. [ISBN: 0-13-332304-8] Refreshing treatment of traditional fare. Emphasis on functions, graphs and problems of the type encountered in calculus. Touches on notions of limit, continuity, average rate of change, linear interpolation and max/min problems. Trigonometry is done analytically. Plenty of exercises, some for "enthusiasts." JK

Education, S(17-18), P. Studies in Mathematics Education, Volume 1. Ed: Robert Morris. UNESCO, 1980, 18 F (P). [ISBN: 92-3-101779-9] First of a four-volume set examining current or emerging programs in mathematics education in different countries. This volume describes mathematics programs in Hungary, Indonesia, Japan, the Philippines, U.S.S.R., the United Kingdom, and the United Republic of Tanzania. Common themes are introduction or modification of a "new math" curriculum and improvement in teacher preparation. MW

Education, P. Visible Language: Understanding the Symbolism of Mathematics. Ed: Richard R. Skemp. Visible Language (Box 1972 CMA, Cleveland, OH 44106), 1982, 103 pp, (P). A special issue of the journal Visible Language (Vol. XVI, No. 3, Summer 1982) devoted to linguistic, psychological and phenomonological analyses of the role played by symbols in elementary mathematics (mostly arithmetic). Offers useful and innovative perspectives on problems faced by children in learning arithmetic. LAS

History, P. Il Circolo Matematico di Palermo. Aldo Brigaglia, Guido Masotto. Edizioni Dedalo, 1982, 443 pp, (P).

History, S(16-18), P, L***.** Zermelo's Axiom of Choice: Its Origins, Development, and Influence. Gregory H. Moore. Stud. in History of Math. & Phy. Sci., No. 8. Springer-Verlag, 1982, xiv + 410 pp, $38. [ISBN: 0-387-90670-3] A fascinating and detailed history of the controversies surrounding the axiom which "epitomizes the fundamental changes--mathematical, philosophical, and psychological --that took place when mathematicians seriously began to study infinite collections of sets." GHM

History, P, L. P.R. Halmos: Selecta--Research Contributions. Ed: Donald E. Sarason, Nathaniel A. Friedman. Springer-Verlag, 1983, xxviii + 458 pp, $32. [ISBN: 0-387-90755-6] A selection of the mathematical writings of Halmos--from 1939 to 1949 on ergodic theory, since then on operator theory--introduced by survey essays on each theme by the two editors. LAS

Foundations, P. Wittgenstein on Rules and Private Language: An Elementary Exposition. Saul A. Kripke. Harvard U Pr, 1982, x + 150 pp, $12.50. [ISBN: 0-674-95400-9] Author contends that Wittgenstein's critical analysis of the paradox of the notion of following a rule played a crucial role in his philosophy of language in general and in his later philosophy of mathematics in

particular. This book presents Kripke's interpretation of Wittgenstein's argument. GHM

Foundations, P. Projections of Lawless Sequences. G.F. van der Hoeven. Math. Centre Tracts, No. 152. Math Centrum, 1982, ii + 237 pp, Dfl. 30,45 (P). [ISBN: 90-6196-244-7] Doctor's thesis on the intuitionistic Baire-space. GHM

Combinatorics, S(18), P. Theory and Practice of Combinatorics. Ed: Alexander Rosa, Gert Sabidussi, Jean Turgeon. Math. Stud., No. 60. Elsevier North-Holland, 1982, x + 263 pp, $59 (P). [ISBN: 0-444-86318-4] A collection of articles honoring Anton Kotzig on the occasion of his sixtieth birthday. The themes of the articles are all somehow related to Professor Kotzig's own recent work. The richness and diversity of his work is reflected in this volume. CEC

Number Theory, S(18), P. Modular Units. Daniel S. Kubert, Serge Lang. Grund. der math. Wissenschaften, B. 244. Springer-Verlag, 1981, xiii + 358 pp, $38. [ISBN: 0-387-90517-0] The basic theory of the units and cuspidal divisor class group in the modular function fields appears in this book. This theory has been developed extensively over the past few years. Includes an extensive list of references. CEC

Algebra, P. Graded Ring Theory. C. Nastasescu, F. van Oystaeyen. Math. Lib., V. 28. Elsevier North-Holland, 1982, ix + 340 pp, $38. [ISBN: 0-444-86489-x] A readable presentation of the major results on graded rings. Important topics discussed include arithmetic, Noetherian and non-commutative graded rings, filtered rings and modules, and homological properties. Many references and exercises given. SG

Calculus, T(16-17: 1), S, P. Vektor- und Tensorrechnung für Ingenieure. R. de Boer. Hochschultext. Springer-Verlag, 1982, ix + 260 pp, $15.90 (P). [ISBN: 0-387-11834-9]

Complex Analysis, S(18), P. Clifford Analysis. F. Brackx, R. Delanghe, F. Sommen. Research Notes in Math., No. 76. Pitman Pub, 1982, 308 pp, $19.95 (P). [ISBN: 0-273-08535-2] This research note presents a function theory in higher dimensions, with applications. This involves a generalization of the classical theory of holomorphic functions of one complex variable, and a refinement of the theory of harmonic functions. This theory is applied to the theory of distributions, harmonic analysis on the unit sphere and transform analysis in Euclidean space. CEC

Differential Equations, P. Generalized Solutions of Hamilton-Jacobi Equations. P.L. Lions. Research Notes in Math., No. 69. Pitman Pub, 1982, 317 pp, $24.95 (P). [ISBN: 0-273-08556-5] Existence and uniqueness theory for solutions to the Cauchy and Dirichlet problems for Hamilton-Jacobi equations. Properties of solutions--"generalized" in the sense that they are locally Lipschitz--are investigated. First section on classical topics is followed by two more on mainly new results. PZ

Differential Equations, T(17: 1), S, P, L. Green's Functions, Second Edition. G.F. Roach. Cambridge U Pr, 1982, xiv + 325 pp, $14.95 (P); $39.50. [ISBN: 0-531-28288-8; 0-521-23890-0] A self-contained introduction to solving boundary-value problems by using Green's functions. This edition is extended with two new chapters and four appendices. The new chapters demonstrate additional ways of calculating the functions and the use of approximate Green's functions. Includes exercises and a bibliography. (First Edition, TR, January 1971.) CEC

Differential Equations, P. Stochastic Differential Equations on Manifolds. K.D. Elworthy. London Math. Soc. Lect. Note Ser., No. 70. Cambridge U Pr, 1982, 326 pp, $27.50 (P). [ISBN: 0-521-28767-7] These very articulate notes represent a much expanded and revised version of notes published by the author and J. Eells in 1974. Much background material, including a very quick review of manifold theory, is provided. JAS

Differential Equations, L. Spectral Transform and Solitons: Tools to Solve and Investigate Non-linear Evolution Equations, Volume One. Francesco Calogero, Antonio Degasperis. Stud. in Math. & Its Applic., V. 13. Elsevier North-Holland, 1982, xv + 516 pp, $111.50. [ISBN: 0-444-86368-0] The spectral transform method can be used to solve certain classes of nonlinear partial differential equations. This introduction to the technique is accessible to readers with only a limited background in the theory of ordinary and partial differential equations. AO

Differential Equations, T(15-16: 1, 2), S*, P, L.** Differential Equation Models. Ed: Martin Braun, Courtney S. Coleman, Donald A. Drew. Modules in Appl. Math., V. 1. Springer-Verlag, 1983, xix + 380 pp, $28. [ISBN: 0-387-90695-9] The first of four volumes of modules containing realistic applications of undergraduate mathematics, drawn from MAA/CUPM summer workshops during the years 1972-1977. This volume on differential equations includes such diverse areas as detection of diabetes, traffic models, biological cycles, and heat transfer. LAS

Functional Analysis, T(18: 1), S, P. Lecture Notes in Mathematics-922: Weak Continuity and Weak Lower Semicontinuity of Non-Linear Functionals. Bernard Dacorogna. Springer-Verlag, 1982, 120 pp, $9.80 (P). [ISBN: 0-387-11488-2] Lecture notes from a Brown University graduate course. Approximate solutions to nonlinear partial differential equations and variational problems may converge weakly to a limit; whether the limit is a solution or "generalized solution" depends on weak continuity or weak semicontinuity of the operator involved. Includes physical applications to nonlinear conservation and elasticity. PZ

Functional Analysis, P. Fourier Series: A Modern Introduction, Volume 2, Second Edition. R.E. Edwards. Grad. Texts in Math., No. 85. Springer-Verlag, 1982, xi + 369 pp, $39.80. [ISBN: 0-387-90651-7] Reflects advances since the 1967 edition (TR, August-September 1968), brings the bibliography up-to-date, still provides a substantial introduction to those facets of Fourier theory "that fit most naturally into function-analytic garb." AWR

Functional Analysis, T(18: 1, 2), S, P. Introduction to Banach Spaces and their Geometry. Bernard Beauzamy. Math. Stud., No. 68. Elsevier North-Holland, 1982, xi + 308 pp, $41.75 (P). [ISBN: 0-444-86416-4] A readable, self-contained intermediate-level graduate text; assumes only elementary real analysis. Develops functional analysis results only as needed in Banach space setting, not in full generality. After preliminaries, sections cover common examples and their geometry, metric properties, geometry of super-reflexive spaces. Exercises follow each chapter. PZ

Functional Analysis, P. Approximation of Hilbert Space Operators, Volume I. Domingo A. Herrero. Res. Notes in Math., No. 72. Pitman Pub, 1982, xiii + 255 pp, $19.95 (P). [ISBN: 0-273-08579-4] Approximations related to similarity invariant operators; norm closure characterizations; estimates of the distance from an operator to a subspace; and compact perturbations. RWN

Analysis, T, S(16-18), P, L. Asymptotic Methods in Analysis. N.G. de Bruijn. Dover Pub, 1981, xii + 200 pp, $4.25 (P). [ISBN: 0-486-64221-6] An unabridged and corrected republication of the Third Edition (1970) of the work originally published in 1958. Attention is focused mainly on methods (especially the saddle point method and iteration methods). Examples, chosen mainly for purposes of instruction, are explained in detail and can be easily followed by anyone with a certain maturity with respect to analysis and some general knowledge of complex function theory. LCL

Analysis, P. Lecture Notes in Mathematics-881: Nonstandard Analysis. Robert Lutz, Michel Goze. Springer-Verlag, 1981, xiv + 261 pp, $16.80 (P). [ISBN: 0-387-10879-3] This work presents a number of examples illustrating the use of nonstandard analysis in the treatment of perturbation problems in algebra and differential equations. It also includes a tutorial introduction to nonstandard analysis and examples of its use in topology and differential calculus. AO

Algebraic Geometry, P. Lectures on p-adic Differential Equations. Bernard Dwork. Grund. der math. Wissenschaften, B. 253. Springer-Verlag, 1982, viii + 310 pp, $46. [ISBN: 0-387-90714-9] A study of p-adic properties of solution to the p-adic hypergeometric differential equation by constructing the so-called associated Frobenius structure. SG

Geometry, T(15-16: 1, 2), L. Geometry: A Metric Approach with Models. Richard S. Millman, George D. Parker. Undergrad. Texts in Math. Springer-Verlag, 1981, x + 355 pp, $29.80. [ISBN: 0-387-90610-X] Intended for a first rigorous course in geometry, this text is based on Birkhoff's metric approach. Six chapters on absolute geometry are followed by chapters on the theory of parallels, hyperbolic geometry, classical Euclidean geometry, area, isometries. Models are used to illustrate axioms, definitions and theorems. Numerous problems. JNC

Geometry, T*(15: 1), S, P*, L**.** Geometrical Investigations: Illustrating the Art of Discovery in the Mathematical Field. John Pottage. Addison-Wesley, 1983, xxi + 480 pp, $35.95. [ISBN: 0-201-05733-6] Scholarly work on the art of mathematical discovery via a sequence of problems in elementary geometry. Cast in the form of a Galilean dialogue. For the non-specialist as well as the specialist. Meticulously-executed drawings. Extensive "Annotationes." Well-graded complementary "Problemata" with "Solutiones." Sure to attract a following. A must buy for every high school, college and university library. JK

Algebraic Topology, P. Lecture Notes in Mathematics-941: Homotopie des Espaces de Sections. André Legrand. Springer-Verlag, 1982, vii + 132 pp, $8.80 (P). [ISBN: 0-387-11575-7]

Optimization, P, L. Lecture Notes in Economics and Mathematical Systems-197: Integer Programming and Related Areas: A Classified Bibliography, 1978-1981. Ed: R. von Randow. Springer-Verlag, 1982, xiv + 338 pp, $27 (P). [ISBN: 0-387-11203-0] The third in a series of bibliographies on integer programming, this volume covers the period from early 1978 to mid 1981. Publications are listed by author and by subject. AO

Probability, T(14-15: 1, 2), S. Probability and Errors for the Physical Sciences. S.K. Muthu. Orient Longman (US Distr: Apt Books), 1982, ix + 568 pp, $35. [ISBN: 0-86131-137-X] A rather expanded treatment (i.e., many examples and extended explanations) of elementary probability and statistics (binomial, Poisson, normal distributions) with a large chapter on curve fitting (the principle of least squares). The language (calculus) and heuristics (motivations and applications) will be comfortable to students in the physical sciences. LCL

Statistics, S(17), P*, L. Residuals and Influence in Regression. R. Dennis Cook, Sanford Weisberg. Chapman & Hall, 1982, x + 230 pp, $25. [ISBN: 0-412-24280-X] Well-written, comprehensive monograph. First half deals with diagnostic methods using residuals, last half with a variety of methods for the study of "influence" (e.g., the effect on the analysis of deleting a case). Methods are amply illustrated on a variety of data sets and with a large number of figures. Also contains a good set of references. RSK

Computer Literacy, T(13-16: 1), S. Computers in Society. Nancy and Robert A. Stern. Prentice-Hall, 1983, xxi + 518 pp, $18.95. [ISBN: 0-13-165282-6] Well organized liberal arts survey. First

half gives overview of history, hardware (including chapter on micros and minis) and software (including chapter on Basic), stressing the user's point of view. Remainder surveys applications and problems in business, education, health, etc., stressing the pros and cons. Coverage is (predictably) superficial in many areas and the bibliography seems unduely limited, but the text is still well worth considering. GHM

Computer Programming, T(13-16: 1). Essentials of Cobol Programming: A Structured Approach. Gerald N. Pitts, Barry L. Bateman. Computer Sci Pr, 1982, ix + 145 pp, $14.95 (P). [ISBN: 0-914894-34-X] A well-organized introductory programming text. It presents the syntax without ever muddying the waters with general ideas from computer science, e.g., data structures or sorting algorithms. The structured style of programming is emphasized. JAS

Computer Programming, T(13: 1), S, P, L. Pascal, An Introduction to Methodical Programming, Second Edition. William Findlay, David A. Watt. Computer Sci Pr, 1981, x + 404 pp, $13.95 (P). [ISBN: 0-914894-73-0] A well-written and easy to follow introduction to structured programming and Pascal. This edition contains more examples and exercises and is now compatible with the proposed ISO standard for Pascal. (First Edition, TR, June-July 1979.) CEC

Computer Programming, S. An Introduction to Visicalc Matrixing for Apple and IBM. Harry Anbarlian. McGraw-Hill, 1982, xiii + 252 pp, $22.95 (P). [ISBN: 0-07-001605-4] This book is primarily for Apple and IBM Personal Computer users. Visicalc matrixing is presented primarily through examples. CEC

Computer Programming, S, P. Basic Programs for Scientists and Engineers. Alan R. Miller. SYBEX, 1981, xvii + 318 pp, $14.95 (P). [ISBN: 0-89588-073-3] The second in the SYBEX Programs for Scientists and Engineers series, this volume includes basic programs for the mean and standard deviation, random numbers, vector and matrix operations, simultaneous solution of linear equations, curve fitting, sorting, Newton's method and numerical integration. JNC

Computer Programming, L. Osborne CP/M User Guide, Second Edition. Thom Hogan. Osborne/McGraw-Hill, 1982, xii + 286 pp, $15.95 (P). [ISBN: 0-931988-82-9] This excellent new edition adds a complete discussion of CP/M-86 for the newer 8088 and 8086 based microcomputers. MP/M II and CP/NET are also discussed. The coverage is very broad--including languages and major applications packages. The coverage is also conservative--avoiding the CP/M users group and other hobbiest support. (First Edition, TR, April 1982.) JAS

Computer Programming, T(15-17), S, P, L. The Science of Programming. David Gries. Texts & Mono. in Comp. Sci. Springer-Verlag, 1981, xiii + 366 pp, $19.80. [ISBN: 0-387-90641-X] This pioneering text shows many signs of hasty composition (e.g., reference to nonexistent exercises, material of questionable relevance, etc.) but brings us closer to the goal of making the composition of clear and correct programs a science rather than an art. Shows how to use predicate calculus to express correctness assertions at the design stage and use them to guide development so that one can rigorously prove correctness of the final program. Many interesting exercises, historical notes. GHM

Software Systems, P, L*. Relational Database Systems: Analysis and Comparison. Ed: Joachim W. Schmidt, Michael L. Brodie. Springer-Verlag, 1983, xv + 618 pp, $19.80. [ISBN: 0-387-12032-7] This volume is a collection of working documents prepared for a group investigating the justifiability of an ANSI standard for relational database management systems. It includes a feature catalogue and detailed analyses and comparisons of fourteen database management systems supporting the relational data model. AO

Software Systems, S, P, L. UNIX Time-Sharing System: UNIX Programmer's Manual, Revised and Expanded Version. Holt, Rinehart & Winston, 1983, xiv + 425 pp, (P). [ISBN: 0-03-061742-1] The complete reference manual for Bell Labs' UNIX operating system: commands, system calls, subroutines, and file information. This is not an introduction to UNIX, but a concise, often cryptic set of reference pages for each command. Local adaptations of UNIX may not conform exactly to these reference pages, but the major features should be available on any UNIX system. LAS

Computer Science, S(15-17), L*. Programming Languages: A Grand Tour. Ellis Horowitz. Computer Sci Pr, 1983, ix + 664 pp, $37.95. [ISBN: 0-914894-67-6] A collection of 24 reprints of important articles on programming languages, some general articles, many on features of particular languages. Sectioned into history and design, the Algol family, applicative languages, data abstraction, concurrency, and languages for the '80's. RWN

Computer Science, T(14-17: 1, 2), S, L. Laboratory Minicomputing. John R. Bourne. Academic Pr, 1981, x + 297 pp, $27. [ISBN: 0-12-119080-3] Presents background material for successful use of minicomputers for activities such as acquiring analog data, displaying data, controlling real time processes, etc. Describes PDP-11 and LSI-11 for hardware and UNIX and C for software. As well, there are chapters on laboratory I/O, interrupts, real time programming. Three appendices. Chapter exercises. References. Index. RJA

Computer Science, S(13-8). Operating Systems: Concepts and Principles. John Zarrella. Microcomputer Applic, 1979, 144 pp, $8.95 (P). [ISBN: 0-935230-00-9] Written for persons who know how to program but are not experts nor intend to become operating systems specialists. Very readable; key new terms on each page in boldface type. Many simple but useful diagrams. Ideal place for an

average microcomputer user to begin to understand how operating systems are organized. Appendices. References. Index. RJA

Computer Science, T(14-15: 1), S, L*. An Introduction to Microcomputers, Volume 1: Basic Concepts, Second Edition. Adam Osborne. Osborne/McGraw-Hill, 1980, x + 433 pp, $12.99 (P). [ISBN: 0-931988-34-9] This is a very fine technical introduction to machine organization in the framework of micro-computers. Its principal weaknesses as a text are the lack of problems and more emphasis on hardware and details like the nature of bisync communication than on elementary algorithms, data structures and programming. The book is extremely useful as a reference and introduction because of its thorough coverage and good index. JAS

Computer Science, S(15-18), P. The HP-IL System: An Introductory Guide to the Hewlett-Packard Interface Loop. Gerry Kane, Steve Harper, David Ushijima. Osborne/McGraw-Hill, 1982, 106 pp, $16.99 (P). [ISBN: 0-931988-77-2] A description of the hardware and software aspects of a new two-wire closed loop interface system proposed by Hewlett-Packard for inexpensive low-power communications (up to 300 feet) suitable for personal computers, laboratory instruments and pocket calculators. Given the sales pitch which HP includes, the price seems a bit high. JAS

Computer Science, P. Calvin C. Elgot: Selected Papers. Calvin C. Elgot. Ed: Stephen L. Bloom. Springer-Verlag, 1982, xxiv + 460 pp, $32. [ISBN: 0-387-90698-3] Photographic reproductions of 13 original papers. Elgot made influential contributions to the study of computation (especially syntax and semantics of flowchart schemes) in the framework of category theory and algebra. GHM

Computer Science, S, P, L. Selected Writings on Computing: A Personal Perspective. Edsger W. Dijkstra. Texts & Mono. in Comp. Sci. Springer-Verlag, 1982, xvii + 362 pp, $28. [ISBN: 0-387-90652-5] Selections from Dijkstra's scientific correspondence during the intellectual preoccupations of one of the pioneers of computer science. Most interesting are numerous "trip reports" (with some names changed to variables to protect against slander suits) in which Dijkstra reports to his employer (Burroughs) on his reactions to visits and lectures in computer facilities around the world. LAS

Applications (Digital Electronics), S(15-16), P. PET/CBM and the IEEE 488 Bus (GPIB), Second Edition. Eugene Fisher, C.W. Jensen. Osborne/McGraw-Hill, 1982, xii + 319 pp, $15.99 (P). [ISBN: 0-931988-78-0] Designed to serve three levels of would-be users of PET/CBM computers in conjunction with laboratory instruments. First comes a cookbook approach; when that fails to work there is a "why" level; and finally full machine level information on the IEEE 488 Bus in the PET/CBM computers is provided for those who would solder up their own devices. JAS

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-201: Price Effects in Input-Output Relations: A Theoretical and Empirical Study for the Netherlands 1949-1967. Paul M.C. de Boer. Springer-Verlag, 1982, x + 140 pp, $10.70 (P). [ISBN: 0-387-11550-1] An approach to adapting various input-output economic models for use with monetary values that are commonly available instead of values which are seldom available. AWR

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-203: Compilation of Input-Output Tables. Ed: Jiri V. Skolka. Springer-Verlag, 1982, vii + 307 pp, $21 (P). [ISBN: 0-387-11553-6] Proceedings of a session of the 17th General Conference of the International Association for Research in Income and Wealth, held in Gouvieux, France. Separate papers on procedures used in various countries for developing data for input-output analysis. AWR

Applications (Physics), P. Abstracts of the Workshop Statistical Mechanics, Dynamical Systems and Turbulence. IMA, 1982, 46 pp, (P). First in a series of reports from the University of Minnesota's Institute for Mathematics and Its Applications: summaries with references of 15 papers on the models that bifurcation theory, iterated maps, and dynamical systems provide for turbulent phenomena. LAS

Applications (Social Science), T(15-16: 1, 2), S*, P, L.** Political and Related Models. Ed: Steven J. Brams, William F. Lucas, Philip D. Straffin, Jr. Modules in Appl. Math., V. 2. Springer-Verlag, 1983, xx + 396 pp, $28. [ISBN: 0-387-90696-7] More applications of undergraduate mathematics--the second of four volumes from the 1972-1977 MAA/CUPM summer workshops. Topics in this volume range from coalition values to apportionment, from indices of voting power to waste water management. An excellent resource for undergraduate seminars in mathematical modelling. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

1983 Officers and Committees

Mathematical Association of America
Officers and Committees
March 15, 1983

General Offices: Dolciani Mathematical Center
1529 Eighteenth Street, N.W., Washington, DC 20036
Telephone 202-387-5200

Executive Director: Alfred B. Willcox
Associate Director: Marcia P. Sward
Editorial Director: Raoul Hailpern
Secretary Emeritus: Henry L. Alder

Officers

President, Ivan Niven, University of Oregon (1983-84)
Past-President, Richard D. Anderson, Louisiana State University (1983)
First Vice-President, Ronald L. Graham, Bell Labs (1982-83)
Second Vice-President, Donald J. Albers, Menlo College (1983-84)
Editor, Paul R. Halmos, Indiana University (1982-1986)
Secretary, David P. Roselle, Virginia Tech (1980-84)
Treasurer, Leonard Gillman, University of Texas (1983-87)

Additional Members of the Board of Governors

Past-Presidents: Henry L. Alder, University of California at Davis (1980-84); Dorothy L. Bernstein (1982-86)
Elected Members of the Finance Committee: G. Baley Price (1980-83) and Donald L. Kreider (1982-85)
Editor of Mathematics Magazine: Doris J. Schattschneider, Moravian College (1981-85)
Editor of American Mathematical Monthly: Paul Halmos (1981-85)
Editor of Two-Year College Mathematics Journal: Warren Page (1983-88)

Governors-at-Large:

Herbert S. Wilf, University of Pennsylvania (1981-83)
Martha Zelinka, Weston High School (1981-83)
Peter L. Renz, W.H. Freeman, San Francisco (1982-1984)
Marilyn J. Zweng, University of Iowa (1982-84)
Eleanor Green Jones, Norfolk State University (1983-85)
Andrew C-F Liu, University of Alberta (1983-85)

Sectional Governors (July 1, 1980-June 30, 1983):

Florida, Beverly L. Brechner, University of Florida
Illinois, Gordon D. Mock, Western Illinois University
Intermountain, C. Edmund Burgess, University of Utah
Iowa, William L. Waltmann, Wartburg College
Louisiana-Mississippi, David E. Cook, University of Mississippi
Maryland-DC-Virginia, Ronald M. Davis, Northern Virginia Community College
Michigan, Delia Koo, Eastern Michigan University
North Central, Dale E. Varberg, Hamline University
Eastern Pennsylvania-Delaware, Gerald J. Porter, University of Pennsylvania
Southern California, John Todd, California Tech
Texas, John T. Mohat, North Texas State University

Sectional Governors (July 1, 1981-June 30, 1984):

Allegheny Mountain, Charles A. Cable, Allegheny College
Indiana, Meyer Jerison, Purdue University
Kentucky, Martha F. Watson, Western Kentucky University
Metropolitan New York, Donald J. McCarthy, St. John's University
Nebraska, Gerald W. Johnson, University of Nebraska
Northern California, Jean J. Pedersen, University of Santa Clara
Oklahoma-Arkansas, James H. Yates, Central State University
Rocky Mountain, John H. Hodges, University of Colorado
Wisconsin, Norbert J. Kuenzi, University of Wisconsin, Oshkosh

Sectional Governors (July 1, 1982-June 30, 1985):

Kansas, Sister Jo Ann Fellin, Benedictine College
Missouri, H. Keith Stumpff, Central Missouri University
New Jersey, Richard F. Gabriel, Seton Hall University
Northeastern, Donald B. Small, Colby College
Ohio, William H. Beyer, University of Akron
Pacific Northwest, Calvin T. Long, Washington State University
Seaway, Violet H. Larney, SUNY at Albany
Southeastern, Allan B. Gray, Jr., Northern Arizona University
Southwestern, Ivey C. Gentry, Wake Forest University

MAA Section Officers

Allegheny Mountain:

Chairman: Franz X. Hiergeist, West Virginia University, Morgantown
First Vice-Chairman: Kathleen A. Taylor, Duquesne University
Second Vice-Chairman: Christine R. Cancro, Penn State, New Kensington
Coordinator of Student Programs: Ronald E. Harrell, Allegheny College
AHSME Regional Exam Coordinators: Nicholas Ford, Pennsylvania State University, Fayette (W. PA); I. Dee Peters, University of West Virginia (WV)
Newsletter Editor: Richard Reese, Edinboro State College
Secretary/Treasurer: John M. Atkins, West Virginia University

Eastern Pennsylvania and Delaware:

Chairman: Bing Wong, Wilkes College
Vice-Chairman: David E. Flesner, Gettysburg College
AHSME Regional Exam Coordinators: Albert E. Filano, West Chester State College; Willard E. Baxter, University of Delaware
Newsletter Editor: Paul Cochrane, Bloomsburg

State College
Secretary/Treasurer: James O. Brooks, Villanova University

Florida:

Chairman: Charles W. Nelson, University of Florida
Chairman-Elect: Don R. Lichtenberg, University of South Florida
Vice-Chairman: Bettye Anne Case, Tallahassee Community College
Vice-Chairman: Roy C. Jones, Jr., University of Central Florida
Past Chairman: Robert Gilmer, Florida State University
AHSME Regional Exam Coordinators: Roy Bolduc, University of Florida; Walter E. Mientka, University of Nebraska (Canal Zone); Eugene A. Francis, University of Puerto Rico
Public Relations: Ralph D. McWilliams, Florida State University
Newsletter Editor: Bruce H. Edwards, University of Florida
Secretary/Treasurer: Ralph D. McWilliams, Florida State University

Illinois:

Chairman: Richard F. Johnsonbaugh, DePaul University
Chairman-Elect: Ronald M. Shelton, Milliken University
First Vice-Chairman: Linda R. Sons, Northern Illinois University
Second Vice-Chairman: John W. Leduc, Eastern Illinois University
Past-Chairman: John S. Haverhals, Bradley University
AHSME Regional Exam Coordinator: Gary Tippett, Bradley University
Public Relations: Claire Krukenberg, Eastern Illinois University
Secretary/Treasurer: Howard C. Saar, 216 Willow Dr. E., Plainfield, IL

Indiana:

Chairman: Carl C. Cowen, Jr., Purdue University
Vice-Chairman: Paul M. Harms, Taylor University
AHSME Regional Exam Coordinator: Stephen Terry, The Nyhart Company
Public Relations: Roger Lautzenheiser, Rose-Hulman Inst of Tech
Newsletter Editor: Roger Lautzenheiser, Rose-Hulman Inst of Tech
Secretary/Treasurer: Richard R. Patterson, Indiana-Purdue University

Intermountain:

Chairman: Donald W. Robinson, Brigham Young University
First Vice-Chairman: Dewey F. Furness, Ricks College
Second Vice-Chairman: Linda C. Hill, Idaho State University
AHSME Regional Exam Coordinators: Ralph V. Skarda, Brigham Young University (UT); Tom Williams, Idaho State University
Secretary/Treasurer: Lawrence O. Cannon, Utah State University

Iowa:

Chairman: Edward T. Hill, Cornell College
Chairman-Elect: Lynn J. Olson, Wartburg College
AHSME Regional Exam Coordinator: Robert Buckingham, Lutheran Mutual Life Insurance Company
Public Relations: James L. Cornette, Iowa State University
Secretary/Treasurer: James Peake, Iowa State University

Kansas:

Chairperson: John W. Carlson, Emporia State University
Chairperson-Elect: Mary Kay Corbitt, University of Kansas
Associate Chairperson for Community Colleges: Jean Moran, Donnelly College
AHSME Regional Exam Coordinator: Steven L. Cooper, Security Benefit Group
Secretary/Treasurer: Robert H. Thompson, Washburn University

Kentucky:

Chairman: Richard M. Davitt, University of Louisville
Chairman-Elect: Peter E. Moore, Northern Kentucky University
Vice-Chairman: Edgar E. Enochs, University of Kentucky
AHSME Regionl Exam Coordinator: William Parker, Jefferson Community College
Newsletter Editor: Glenn E. Johnston, Morehead State University
Secretary/Treasurer: Joe K. Smith, Northern Kentucky University

Louisiana-Mississippi:

Chairman: Paul W. Spikes, Mississippi State University
LA Vice-Chairman: Kathleen A. Drude, Northeast Louisiana University
MS Vice-Chairman: Ken Johnston, Hinds Junior College
Past Chairman: Adam J. Hulin, University of New Orleans
AHSME Regional Exam Coordinators: James E. Keisler, Louisiana State University; David E. Cook, University of Mississippi
Newsletter Editor: John L. Tilley, Mississippi State University
Secretary/Treasurer: John L. Tilley, Mississippi State University

Maryland-DC-Virginia:

Chairman: John Schmeelk, Virginia Commonwealth University
Chairman-Elect: Howard L. Penn, U.S. Naval Academy
Vice-Chairman for Programs: Robert E. Lewand, Goucher College
Treasurer: John Schmeelk, Virginia Commonwealth University
AHSME Regional Exam Coordinator: Edward D. Bender, J. Sargeant Reynolds Community College
Public Relations: Caren L. Diefenderfer, Hollins College
Secretary: Caren L. Diefenderfer, Hollins College

Metropolitan New York:

Chairman: Theresa J. Barz, St. John's University
Vice-Chairman for Four-Year Colleges: John A. Thorpe, SUNY at Stony Brook

Vice-Chairman for Two-Year Colleges: Douglas S. Brown, Nassau Community College
Vice-Chairman for High Schools: Cecile M. Cohen, Horace Mann School
Treasurer: Robert G. Payton, Adelphi University
AHSME Regional Exam Coordinator: Sandra Pulver, Pace University
Newsletter Editor: William F. Orr, Hofstra University
Secretary: Lily E. Christ, John Jay College of Criminal Justice

Michigan:

Chairman: M.S. Ramanujan, University of Michigan
Vice-Chairman: George Van Zwalenberg, Calvin College
Vice-Chairman for Two-Year Colleges: John S. Kostoff, Delta College
AHSME Regional Exam Coordinator: Joanne Rankin, Eastern Michigan University
Public Relations: Thomas Elsner, General Motors Institute
Newsletter Editor: Yousef Alavi, Western Michigan University
Secretary/Treasurer: Douglas W. Nance, Central Michigan University

Missouri:

President: Kenneth W. Lee, Missouri Western State College
Vice-President: Victor H. Gummersheimer, Southeast Missouri State University
Chairman of High School Lecturers: Leonard Palmer, Southeast Missouri State University
Past President: A. Glen Haddock, University of Missouri
AHSME Regional Exam Coordintor: Alvin Tinsley, Central Missouri State University
Newsletter Editor: Ed W. Huffman, Southwest Missouri State University
Secretary/Treasurer: Jerry Wilkerson, Missouri Western State College

Nebraska:

Chairman: John K. Karlof, University of Nebraska
Chairman-Elect: Allan C. Peterson, University of Nebraska
AHSME Regional Exam Coordinator: Richard Barlow, Kearney State College
Public Relations: H. M. Cox, 1145 N. 44th St., Lincoln, NE
Newsletter Editor: H. M. Cox & T. S. Shores, University of Nebraska, Lincoln
Secretary/Treasurer: Henry M. Cox, 1145 N. 44th St., Lincoln, NE

New Jersey:

Chairman: Susan G. Marchand, Kean College
Chairman-Elect: S. Ashby Foote, Rutgers University
Vice-Chairman for Teaching Innovations: Charles J. Lewis, Monmouth College
Vice-Chairman for Speakers: Sr. Stephanie Sloyan, Georgian Court College
Vice-Chairman for Two-Year Colleges: James J. Magliano, Union College
Vice-Chairman for High Schools: Francis E. Masat, Glassboro State College
Treasurer: Judith B. Seery, Bell Telephone Labs, Murray Hill
AHSME Regional Exam Coordinator: Robert Mitchell, Glassboro State College
Newsletter Editor: Susan G. Marchand, Kean College
Secretary: F. Eugene Clark, 820 Shadowlawn Drive, Westfield, NJ

North Central:

President: Sabra S. Anderson, University of Minnesota
President-Elect: Murray Braden, Macalester College
Member at Large: James Hewitt, Inver Hills Community College
Member at Large: James S. Rue, University of North Dakota
Past President: Roger B. Kirchner, Carleton College
AHSME Regional Exam Coordinators: Wayne Roberts, Macalester College; John Whitcomb, University of North Dakota; John Barsby, St. John's Ravenscourt (Manitoba); Don Kapoor, University of Regina (Saskatchewan); Jan Vandever, South Dakota State University
Public Relations: Joseph Gallian, University of Minnesota
Newsletter Editor: Gerald E. Lenz, St. John's University
Secretary/Treasurer: Allan M. Kirch, Macalester College

Northeastern:

Chairman: James E. Ward, Bowdoin College
Vice-Chairman: Eric C. Nummela, New England College
Two-Year College Rep: Nancy Myers, 67 Highland Ave., Cambridge, MA
AHSME Regional Exam Coordinators: David Rogers, CIGNA, Hartford, CT; Olin Sawyer, Union Mutual Life Insurance Company; Murray Abramson, Bridgewater State College; Owen R. Koppang, Concord High School; J.S.H. Johansson, University of Vermont; William Crawford, Mt. Allison University (Newfoundland, New Brunswick, Nova Scotia, PE Island)
Newsletter Editor: Eric C. Nummela, New England College
Secretary/Treasurer: Gordon D. Prichett, Babson College

Northern California:

Chairman: Roy W. Ryden, Humboldt State University
Vice-Chairman: Michael Thibodeaux, Menlo College
Program Chairman: Hans Samelson, Stanford University
AHSME Regional Exam Coordinators: Robert McFarland, Alameda County Office of Education; Jack Johnson, Brigham Young University (HI, Pacific Islands, Guam)
Newsletter Editor: William G. Chinn, 539 29th Ave., San Francisco, CA
Secretary/Treasurer: Leonard F. Klosinski, University of Santa Clara

Ohio:

Chairman: Darrell J. Horwath, John Carroll University
Chairman-Elect: Richard A. Little, Baldwin-Wallace College
Program Chairman: Joan P. Leitzel, Ohio State University
Two-Year College Rep: Cynthia P. Yang, Miami University

Past Chairman: J. Douglas Faires, Youngstown State University
AHSME Regional Exam Coordinators: Leo J. Schneider, John Carroll University (OH); I. Dee Peters, University of West Virginia
Public Relations: Susan G. Schneider, John Carroll University
Newsletter Editor: David E. Kullman, Miami University
Secretary/Treasurer: Andrew Sterrett, Jr., Denison University

Oklahoma-Arkansas:

Chairman: Allan C. Cochran, University of Arkansas
First Vice-Chairman: Morris L. Marx, University of Oklahoma
Second Vice-Chairman: John W. Watson, Arkansas Tech University
Past Chairman: Robert D. McMillan, Oklahoma Christian College
AHSME Regional Exam Coordinators: Thomas Cairns, University of Tulsa; John Hodges, University of Arkansas, Little Rock
Public Relations: James E. Scroggs, University of Arkansas
Newsletter Editor: Raymond E. Beasley, Central State University
Secretary/Treasurer: John Jobe, Oklahoma State University

Pacific Northwest:

Chairman: Beryl M. Green, Western Oregon State College
Chairman-Elect: Marjorie M. Enneking, Portland State University
Vice-Chairman of Two-Year Colleges: John C. Bradshaw, Cariboo College
Vice-Chairman of Four-Year Colleges: Duane W. DeTemple, Washington State University
AHSME Regional Exam Coordinators: Donval R. Simpson, Institute of Arctic Biology; Tom Williams, Idaho State University; William Myers, University of Montana; Marion Walter, University of Oregon; Calvin Long, Washington State University; George Sands, University of Calgary (Alberta); Leon Bowden, University of Victoria (British Columbia, Yukon, Northwest Territories)
Public Relations: Sue Kaplan, Western Washington University
Newsletter Editor: Sue Kaplan, Western Washington University
Secretary/Treasurer: Sam L. Dunn, Seattle Pacific University

Rocky Mountain:

Chairman: George S. Donovan, Metropolitan State College
Chairman-Elect: Carl M. Kerns, Mesa College
Vice-Chairman: Marie A. Ritten, National College
Program Chairman: Frank R. DeMeyer, Colorado State University
AHSME Regional Exam Coordinators: Bob Vunovich, University of Southern Colorado; William Myers, University of Montana
Newsletter Editor: David W. Ballew, SD School of Mines & Tech
Secretary/Treasurer: David W. Ballew, SD School of Mines & Tech

Seaway:

Chairperson: Kenneth D. Magill, SUNY Center at Buffalo
Chairperson-Elect: Charles L. Smith, SUNY College at Potsdam
First Vice-Chairperson: Kenneth Wooster, SUNY, College at Cortland
Second Vice-Chairperson: Robert C. Knapp, Herkimer County Community College
Past Chairperson: Howard E. Bell, Brock University
AHSME Regional Exam Coordinators: Douglas Cashing, St. Bonaventure University; Louis F. Scholl, Trocaire College; J.R. Vanstone, University of Toronto; Saverio Mirarchi, Laval Catholic High School
Public Relations: Donald Trasher, SUNY College at Geneseo
Newsletter Editor: Donald Trasher, SUNY College at Geneseo
Secretary/Treasurer: Donald Trasher, SUNY College at Geneseo

Southeastern:

Chairman: Mary F. Neff, Emory University
Chairman-Elect: Isaac S. Metts, The Citadel
Vice-Chairman: Sharon C. Ross, Dekalb Community College
Lecturer: John D. Neff, Georgia Institute of Technology
AHSME Regional Exam Coordinators: James O'Neal, Troy State University; Jan List Boal, Georgia State University; Hal B. Phillips, Jr., Jefferson Stand Life Ins Co; Stanley Lukawecki, Clemson University; David Collings, National Life & Accident Ins Co
Newsletter Editor: DeWayne S. Nymann, University of Tennessee
Secretary/Treasurer: Ivey C. Gentry, Wake Forest University

Southern California:

Chairman: Judith Grabiner, 424 West 7th Street, Claremont
First Vice-Chairman: Robert F. Brown, University of California, Los Angeles
Second Vice-Chairman: Donald Y. Goldberg, Occidental College
Program Chairman: Michael J. Hoffman, California State University
Past Chairman: John Greever, Harvey Mudd College
AHSME Regional Exam Coordinator: Dwight O. Coblentz, Department of Ed
Public Relations: William Margulies, California State University
Secretary/Treasurer: William Margulies, California State University

Southwestern:

Chairman: Ralph W. Ball, New Mexico Institute of Mining & Tech
Vice-Chairman: Alvin Swimmer, Arizona State University
AHSME Regional Exam Coordinator: D. R. Arterburn, New Mexico Institute of Mining & Tech
Secretary/Treasurer: Edward D. Gaughan, New Mexico State University

Texas

Chairman: John Ed Allen, North Texas State University
Chairman-Elect: Robert N. Cranford, University of Texas, Tyler

Level I Director: Vivian A. Dennis, Eastfield College
Level II Director: R. Vic Morgan, Sul Ross State University
Level III Director: James N. Younglove, University of Houston
Director at Large: Nancy C. Ogden, Garland Independent School Dist
Arrangements Chair: Melvin R. Hagan, North Texas State University
Arrangements Chair-Elect: Thomas Keagy, University of Texas
Past Chairman: Robert L. Tennison, University of Texas
AHSME Regional Exam Coordinator: James R. Boone, Texas A & M University
Public Relations: James C. Bradford, Abilene Christian University
Newsletter Editor: Glen E. Mattingly, Sam Houston State University
Secretary/Treasurer: Glen E. Mattingly, Sam Houston State University

Wisconsin:

Chairman: Lois B. Brualdi, Edgewood College
Chairman-Elect: Richard L. Poss, St. Norbert College
Past Chairman: Karl Beres, Ripon College
AHSME Regional Exam Coordinator: Arthur Brownell, Jr., University of Wisconsin
Newsletter Editor: Nichols C. Bystrom, Northland College
Secretary/Treasurer: James M. Sobota, University of Wisconsin

Committees of the Association

Terms of members expire, except where otherwise noted, at the Annual Meeting in January following the last year of service listed below. For temporary committees, no terms are listed since they are automatically discharged at the expiration of the President's term of office, which is the Annual Meeting in January, 1984.

Executive Committee:

Ivan Niven, Chairman (1983-85); Leonard Gillman (1983-87); Ronald L. Graham (1982-83); Paul R. Halmos (1982-86); David P. Roselle (1980-84), all are ex-officio.

Finance Committee:

Ivan Niven, Chairman (1983-85), ex-officio; Richard D. Anderson, (1983-84) ex-officio; Leonard Gillman (1983-87), ex-officio; Donald L. Kreider (1978-85); G. Baley Price (1980-83); David P. Roselle (1980-84), ex-officio.

Audit and Budget Committee: G. Baley Price, Chairman (1980-83); Donald L. Kreider, (1982-85) ex-officio.

Investment Committee: Leonard Gillman, Chairman (1983-87), ex-officio; Henry L. Alder (1982-83); Harley Flanders (1982-83); Gerald Porter (1983-85); David P. Roselle (1980-84) ex-officio; James M. Vaughn, Jr. (1981-83).

Staff and Services Committee: Leonard Gillman, Chairman (1983-87); Ivan Niven (1983-85), both ex-officio.

Advisory Committee on the Archives of American Mathematics:

Judith V. Grabiner, Chairman (1979-85); Leonard Gillman (1979-85); Lucille E. Whyburn (1981-83).

Committee on Corporate Members:

Richard A. Moynihan (1981-85); Clayton V. Aucoin (1981-83); Jeremiah J. Lyons (1981-83); Henry O. Pollak (1983-85); Alfred B. Willcox (1980-85), ex-officio.

Committee on Earle Raymond Hedrick Lectures:

George E. Andrews, Chairman (1981-83); James W. Cannon (1983-85); Daniel Gorenstein (1982-84).

Committee on High School Contests:

Stephen B. Maurer, Chairman (1981-83); George Berzsenyi (1983-85); James D. Bristol (1983-85); Thomas R. Butts (1982-84); Michael W. Ecker (1982-84); Samuel L. Greitzer (1981-83); William Hall (1982-84); Mary Jones (1982-83); Irwin W. Kaufman (1982-85); Murray Klamkin (1981-85); Joseph Konhauser (1982-85); Walter E. Mientka (1979-85); Richard S. Pieters (1981-83); Stanley Rabinowitz (1979-84); Leo J. Schneider (1980-85); Margaret Wilkinson (1982-85); Melvin R. Woodard (1981-83).

Subcommittee on the Annual High School Mathematics Examination: Stephen B. Maurer, Chairman (1981-83); George Berzsenyi (1982-85), ex-officio; Thomas R. Butts (1980-84); Mary H. Jones (1982-83); Murray S. Klamkin (1982-85), ex-officio; Walter E. Mientka (1980-85), ex-officio; Richard S. Pieters (1980-83); Leo J. Schneider (1980-85); Melvin R.Woodard (1980-83).

Subcommittee on the American Intermediate Mathematics Examination: George Berzsenyi, Chairman (1982-85); James D. Bristol (1982-85); Michael W. Ecker (1982-84); Anthony M. Gaglione (1980-83); Murray S. Klamkin (1982-85), ex-officio; Stephen B. Maurer (1981-83), ex-officio; Walter E. Mientka (1980-82), ex-officio; Stanley Rabinowitz (1980-84).

Advisory Panel of the Committee on High School Contests: Pamela Drummond (1982-83); Attila Mate (1980-83); Nathan I. Mendelsohn (1981-83); Harold B. Reiter (1980-83); Akehiko Takahashi (1982-83); Eugene Veklerov (1980-83); Dorothy Wendt (1982-83).

Subcommittee on the USA Mathematical Olympiad: Murray S. Klamkin, Chairman (1982-85); Joshua Barlaz (1982-83); George Berzsenyi (1980-85), ex-officio; Samuel L. Greitzer (1982-83); Irwin W. Kaufman (1982-85); Joseph Konhauser (1982-85); Andrew Liu (1981-83); Stephen B. Maurer (1981-83), ex-officio; Walter E. Mientka (1981-85), ex-officio; Cecil C. Rousseau (1982-84).

Committee on Placement Examinations:

Richard H. Prosl, Chairman (1982-84); John W. Kenelly (1982-85); Bernard L. Madison (1982-83); Billy E. Rhoades (1982-85); Billie A. Rice (1982-83); Thomas W. Tucker (1982-84).

Early Testing Program Panel: Bernard L.

Madison, Chairman (1982-83); D. Gulick (1982-83); Bert Waites (1982-83).

Committee on Publications:

Alan C. Tucker, Chairman (1982-85); Donald J. Albers (1979-83), ex-officio; Lida K. Barrett (1981-83); Charles A. Cable (1982-83); Daniel T. Finkbeiner, II (1981-83); Ross L. Finney (1980-85); Leonard Gillman (1983-87), ex-officio; Paul R. Halmos (1982-86), ex-officio; John Karlof (1982-85); William F. Lucas (1982-84); Barbara L. Osofsky (1981-83); Warren Page (1981-83); Doris Schattschneider (1981-85), ex-officio; Ted Vessey (1982-85).

Subcommittee on Basic Library Lists: Daniel T. Finkbeiner, II, Chairman (1981-83); Deborah T. Haimo (1983-86); William F. Lucas (1982-85); Ted Vessey (1983-86).

Subcommittee on Carus Monographs: Barbara L. Osofsky, Chairman (1981-83); Daniel T. Finkbeiner, II (1981-83); John Karlof (1982-85).

Subcommittee on Carl B. Allendoerfer Awards: Thomas W. Tucker, Chairman (1980-83); Joanne Growney (1982-85); Doris W. Schattschneider (1982-85); J.A. Seebach (1982-84).

Subcommittee on Dolciani Mathematical Expositions: Ross A. Honsberger, Chairman (1981-83); Joseph Malkevitch (1981-83); Barbara Osofsky (1981-83); Donald B. Small (1982-85).

Subcommittee on George Polya Awards: Ronald M. Davis, Chairman (1982-83); Donald J. Albers (1982-84), ex-officio; Larry A. Curnutt (1982-85); Edward J. Specht (1982-84).

Subcommittee on Lester R. Ford Awards: Peter L. Duren, Chairman (1981-83); David Gale (1982-84); Paul R. Halmos (1983-86), ex-officio; Richard S. Millman (1982-85).

Subcommittee on MAA Notes: Alan C. Tucker, Chairman (1982-85); Charles A. Cable (1982-84); William F. Lucas (1982-85).

Subcommittee on MAA Studies in Mathematics: Alan C. Tucker, Chairman (1980-85); Donald W. Anderson (1981-83); Warren Page (1982-85), ex-officio; Douglas B. West (1982-84).

Subcommittee on New Mathematical Library: William G. Chinn, Chairman (1982-85); Basil Gordon (1980-85); Anneli Lax, ex-officio; Max M. Schiffer (1979-85); Ted Vessey (1982-85).

Editor of the New Mathematical Library: Anneli Lax.

Committee on Articulation:

Billy E. Rhoades, Chairman (1982-84); Philip C. Curtis, Jr. (1982-84); Ann D. Holley (1982-84); Chancey Jones (1982-84); Joan P. Leitzel (1982-84); Vincent O'Connor (1982-84); Barbara Wickless (1982-84); Gail S. Young (1982-84).

Committee on Mini-Courses:

John O. Riedl, Chairman (1982-84); Frederick Hoffman (1982-84); Warren Page (1982-85); Jean Pedersen (1981-84).

Committee on National Meeting Programs:

C. E. Burgess, Chairman (1982-84); Ronald L. Graham (1981-83); Peter A. Lindstrom (1982-85); David P. Roselle (1980-84), ex-officio.

Committee on Secondary School Lectures:

Donald B. Small, Chairman (1979-83); Edward Z. Andalafte (1980-84); Henry Gore (1982-85); John M. Jobe (1981-83); Amy King (1982-85); Gary J. Sherman (1982-85); David Skoug (1982-85).

Committee on Sections:

Samuel W. Hahn, Chairman (1981-84); Yousef Alavi (1980-83); David W. Ballew (1981-84); Barbara T. Faires (1982-85); Glen E. Johnston (1982-85); Kenneth R. Rebman (1981-84); Alfred B. Willcox, ex-officio.

Committee on Special Funds of the Association:

G. Baley Price, Chairman (1981-83); Lida K. Barrett (1981-83); Leonard Gillman (1980-87), ex-officio; John M. Jobe (1982-84); Eileen Poiani (1982-84); Richard M. Schori (1982-84).

Committee on the Award for Distinguished Service to Mathematics:

Ralph P. Boas, Chairman (1982-83); Peter J. Hilton (1983-85); Eileen L. Poiani (1982-84).

Committee on the Chauvenet Prize:

Lynn Steen, Chairman (1982-83); Peter J. Hilton (1982-85); Gilbert Strang (1982-84).

Committee on the Exchange of Information on Mathematics:

Yousef Alavi, Chairman (1981-83); Donald M. Hill (1981-83); James R. Leitzel (1979-83); Jean J. Pedersen (1981-83); Lynn A. Steen (1979-83); Marion D. Wetzel (1981-83).

Committee on the Membership of the Association:

Richard D. Anderson, Chairman (1983-85); Leonard Gillman (1983-87), ex-officio; Samuel W. Hahn (1982-84); Ivan Niven (1983-84), ex-officio; David P. Roselle (1980-84), ex-officio; Sue H. Whitesides (1981-83).

Committee on the Putnam Prize Competition:

Gerald L. Alexanderson, Co-Chairman (1983-86); Douglas A. Hensley, Co-Chairman (1981-83); Abraham P. Hillman, Associate Director (1978-86); Melvin Hochster (1982-84); Leonard F. Klosinski (1980-86); Bruce A. Reznick (1982-85).

Committee on the Teaching of Undergraduate Mathematics:

Alan H. Schoenfeld, Chairman (1982-85); Donald W. Bushaw (1980-85); Ronald M. Davis (1982-84); Gloria F. Gilmer (1981-83); David I. Schneider (1981-83); Martha J. Siegel (1981-83); James E. Ward, III (1982-84); June P. Wood (1981-83).

Committee on Two-Year Colleges:

John D. Bradburn, Chairman (1981-83); Thomas A. Carnevale (1982-84); Betsy Darden-Smith (1982-

84); Ann D. Holley (1982-84); J.C. Moss (1981-83); Warren Page (1982-85), ex-officio; Jean B. Smith (1982-85); Elaine L. Tatham (1981-83); Etta Whitton (1982-85); William R. Wilson (1981-84).

Committee on the Undergraduate Program in Mathematics:

Jerome A. Goldstein, Chairman (1983-84); Gerald L. Alexanderson (1980-85); Donald W. Bushaw (1982-85); Betty A. Case (1982-84); John B. Conway (1982-85); James A. Donaldson (1982-84); Joan P. Leitzel (1982-84); Stephen B. Maurer (1981-83); Ivan Niven (1983-84), ex-officio; Mary A. Norton (1982-85); Thomas G. Proctor (1982-84); Billy E. Rhoades (1981-83); Alan H. Schoenfeld (1982-85).

Committee on Visiting Lecturers and Consultants:

Theodore S. Chihara, Chairman (1981-83); Vivian Y. Kraines (1981-83); Jon M. Laible (1980-85); Eric C. Nummela (1982-85); Eugene R. Seelbach (1979-84).

Committee on World War II History:

G. Baley Price, Chairman (1981-83); Churchill Eisenhart (1981-83); Carroll V. Newsom (1981-83); Mina S. Rees (1981-83); J.B. Rosser (1981-83); Raymond L. Wilder (1981-83).

Joint Committees

Joint Committee on Employment Opportunities:

Wilfred E. Barnes, Chairman (1984, MAA); James W. Daniel (1981-83, SIAM); Leonard J. Guibas (1985, SIAM); Calvin Long (1981-83, MAA); Donald C. Rung (1981-84, AMS).

Joint Committee on Retraining for Computer Science:

Donald L. Kreider, Chairman; Susan J. Devlin; Terry Fredericks; Pat Goldberg; Zaven Karian; Anthony Ralston; Olaf Stakelberg.

Joint Committee on Women in Mathematics:

Mary W. Gray, Chairman (1981-83); Etta Z. Falconer (1983, AMS); Pamela Cook-Ioannidis (1983, SIAM); Carol B. LaCampågne (1984, MAA); Edith H. Luchins (1983, MAA); Katherine Pedersen (1983, NCTM); Joel Schneider (1983, NCTM); Barbara Searle (1983, NCTM); Gail A. Williams (1983, MAA).

Joint Meetings Committee:

William J. LeVeque, Chairman; Everett Pitcher; David P. Roselle; Alfred B. Willcox, all ex-officio.

CEEB-MAA Joint Committee on Mutual Concerns:

David P. Roselle, Chairman (1984, MAA); Henry Alder (1985, MAA); Chancey O. Jones (CEEB); John W. Kenelly (1985, CEEB); Donald L. Kreider (CEEB); Ivan Niven (1984, MAA).

Editorial Boards of the Association

American Mathematical Monthly (All terms expire December 31, 1986):

Editor: Paul R. Halmos

Associate Editors: Gerald L. Alexanderson, Sabra Anderson, Sheldon Axler, Thomas F. Banchoff, Ralph Boas, David Borwein, W.W. Comfort, Hugh M.W. Edgar, Allan L. Edmonds, John H. Ewing, Richard K. Guy, Raoul Hailpern, Richard S. Millman, Dale H. Mugler, Kenneth R. Rebman, Seymour Schuster, J.A. Seebach, Lynn A. Steen, Mary R. Wardrop, Robert F. Wardrop, Herbert S. Wilf.

Mathematics Magazine (All terms expire December 31, 1985):

Editor: Doris L. Schattschneider

Associate Editors: Edward J. Barbeau, John A. Beidler, Paul J. Campbell, Underwood Dudley, Gerald A. Edgar, Joseph A. Gallian, Judith V. Grabiner, Raoul Hailpern, Pierre J. Malraison, Joseph Malkevitch, William A. McWorter, Jr., Leroy F. Meyers, Jean J. Pedersen, Gordon Raisbeck, Ian Richards, Eric S. Rosenthal, Dan Shapiro, David A. Smith.

Two-Year College Mathematics Journal (All terms expire December 31, 1988):

Editor: Warren Page

Associate Editors: Jeanne Agnew, Gerald L. Alexanderson, John D. Bradburn, Donald W. Bushaw, Michael W. Chamberlain, Leo Chosid, Gino T. Fala, Stanley Friedlander, Samuel A. Greenspan, Erwin Just, Peter L. Renz, Billie A. Rice, Milton E. Winger.

Newsletter Editorial Committee:

Ronald M. Davis, Chairman (1985); Henry L. Alder (1984); Charles A. Cable (1985), ex-officio; Philip M. Cheifitz (1986); Ronald M. Davis (1985); Katherine P. Layton (1986); James R. Leitzel (1985), ex-officio; Stephen B. Maurer (1985), ex-officio; Eileen Poiani (1985); Kenneth R. Rebman (1986); David P. Roselle (1984), ex-officio; J.A. Seebach (1984); Lynn A. Steen (1984); Marcia P. Sward, ex-officio; Alan C. Tucker (1985).

Representatives of the Association

On Sections of AAAS:

Section A: Lowell J. Paige (1981-83)
Section Q: Joan P. Leitzel (1981-83)
Section U: Fred Mosteller (1981-83)
Section X: Alfred B. Willcox (1981-83)

On UMAP Editorial Board: Peter A. Lindstrom (1981-83); Anne F. O'Neill (1981-83); Philip D. Straffin (1982-84).

On the Council of the Conference Board of the Mathematical Sciences: Ivan Niven.

On Discipline Committee for Mathematics of CEEB: Donald L. Kreider.

On the Governing Council of Mu Alpha Theta: Katherine P. Layton (1982-84).

On the U.S. Commission on Mathematical Instruction: Leon A. Henkin (1978-83); Eileen L. Poiani (1981-85).

Throughout this discussion, we have limited our attention to projections onto coordinate axes only. However, it should be clear that any set of d mutually orthogonal directions could serve as the "coordinates" used in the theorem.

In conclusion, there is a continuous version of the theorem. Presumably, it could be proved directly from the integral form of Hölder's Inequality stated as inequality 188 in Hardy, Littlewood, and Pólya [2, p. 140] namely

$$\left(\prod_{j=1}^{p} \int f_j(x)\,dx\right)^{1/p} \geqslant \int \left(\prod_{j=1}^{p} f_j(x)\right)^{1/p} dx.$$

We state the continuous version as a corollary to the finite version using $m_d(X)$ to denote the d-dimensional measure of the set X.

COROLLARY. *If X is a set with finite positive Riemann measure,*

$$\text{Geom. Mean}\,\{m_s(\pi(X,S)) : |S| = s \text{ and } S \subset N_d\} \geqslant m_d(X)^{s/d}.$$

Sketch of proof. We shall replace the infinite set X by a sequence of finite sets X_k for each natural number k. Partition E^d into d-dimensional cubes with edge $1/k$ and measure $1/k^d$. Select the central point of each cube to lie in X_k if and only if the entire cell is contained in X. Now $m_d(X) \geqslant |X_k|/k^d$ and analogously $m_s(\pi(X,S)) \geqslant p(X_k,S)/k^s$. Take the geometric mean over all $S \subset N_d$ of order s and apply the theorem to the projections of X_k to obtain

$$\text{Geom. Mean}\,\{m_s(\pi(X,S)) : |S| = s \text{ and } S \subset N_d\} \geqslant |X_k|^{s/d}/k^s. \tag{17}$$

Since $\lim_{k\to\infty}|X_k|/k^d = m_d(X)$, the limit as $k \to \infty$ of the right-hand side is $m_d(X)^{s/d}$ as required.

Acknowledgment. The first author is grateful to the University of Waterloo and particularly to the Department of Combinatorics and Optimization for the hospitality extended to him during his recent sabbatical from the U.S. Naval Academy. The work of the second author was supported by NSERC grant A-8237.

References

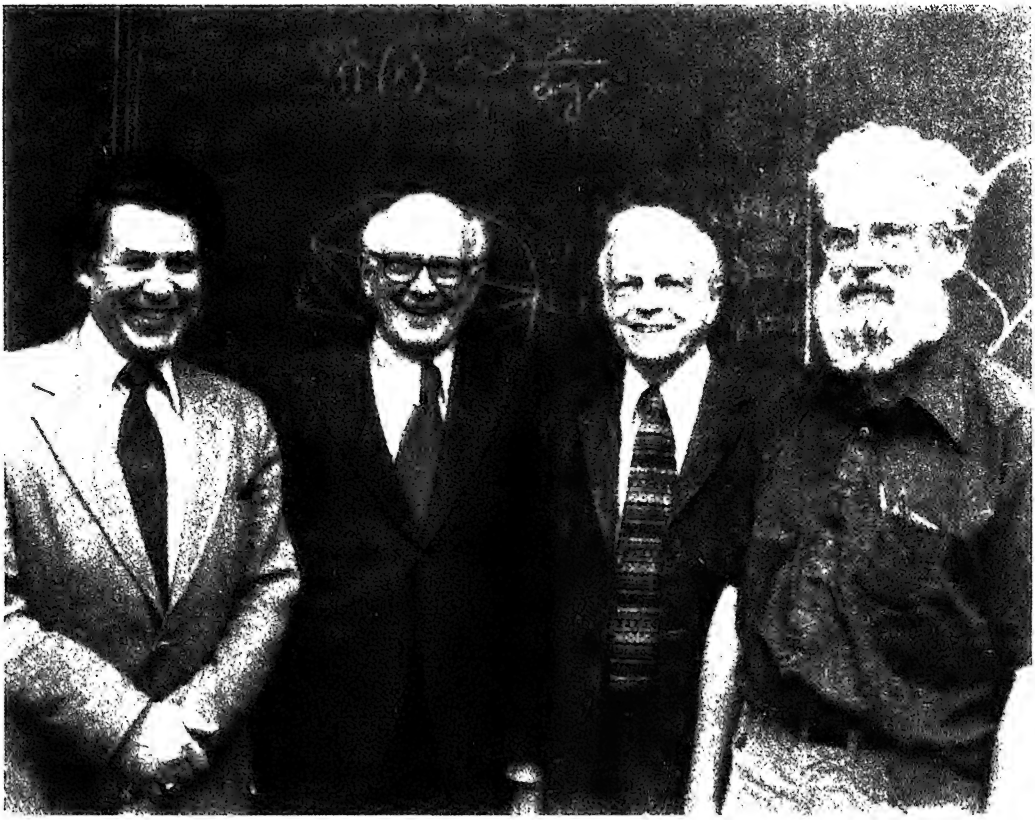
1. H. Alt, K. Mehlhorn, and J. I. Munro, Partial match retrieval in implicit data structures, unpublished manuscript.

2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1967.

MISCELLANEA

103. Among the miscellaneous reading of my youth was a history of modern Europe, which concluded with a general survey and attempted forecast of progress in arts, science, and literature. So far as I can judge, this work was written about the time of Euler or Lagrange. On the subject of mathematics the writer's conclusion was that fruitful investigation seemed at an end, and that there was little prospect of brilliant discoveries in the future. To us, a century later, this judgment might seem to illustrate the danger of prophesying, and lead us to look upon the author as one who must have been too prone to hasty conclusions. I am not sure that careful analysis would not show the author's view to be less rash than it may now appear.

—SIMON NEWCOMB, address to the New York Mathematical Society, Bull. N.Y. Math. Soc., 3(1894) 95–107.



Who are these four heads? See page 340.

$$\sup_{0\leqslant u<2\pi} g(u) = \text{essential}\sup_{0\leqslant u<2\pi} g(u) = \sup_{|z|<1} |F(z)|.$$

Solution by the proposer. Let C_r denote the circle with radius r and center 0. By Cauchy's integral formula, for n a positive integer and $|z| < r < 1$,

$$|F(z)|^n = \left|(2\pi i)^{-1}\int_{C_r} F(v)^n (v-z)^{-1}\,dv\right|$$
$$\leqslant (2\pi)^{-1}(r-|z|)^{-1}\int_0^{2\pi}|F(re^{iu})|^n\,du.$$

So

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1}\liminf_{r\to 1-}\int_0^{2\pi}|F(re^{iu})|^n\,du.$$

Now $|F(re^{iu})|^n \leqslant h(u)^n$ for all $r<1$, and it follows from the Lebesgue dominated convergence theorem that

$$\liminf_{r\to 1-}\int_0^{2\pi}|F(re^{iu})|^n\,du \leqslant \int_0^{2\pi} g(u)^n\,du.$$

Thus

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1}\int_0^{2\pi} g(u)^n\,du \leqslant (1-|z|)^{-1}\Big(\operatorname*{ess.\,sup}_{0\leqslant u<2\pi} g(u)\Big)^n$$

and $|F(z)| \leqslant (1-|z|)^{-1/n}(\text{ess.}\sup_{0\leqslant u<2\pi} g(u))$. Let $n\to\infty$ to obtain $|F(z)| \leqslant \operatorname*{ess.\,sup}_{0\leqslant u<2\pi} g(u)$. The rest is clear.

Tangents to a Circle and a Parabola

6364 [1981, 711]. *Proposed by the late K. B. Leisenring, University of Michigan*

A circle with center at the vertex and radius equal to the latus rectum meets a parabola at P, Q. The circle and parabola have common tangents meeting the parabola at X, Y. Prove that XP, YQ are tangent to the circle.

Solution by Anders Bager, Akelejevej 5, DK-9800 Hjørring, Denmark. The parabola $y^2 = x$ and the circle $x^2+y^2=1$ meet at $P = (a,\sqrt{a})$ and $Q = (a,-\sqrt{a})$ where $a = (\sqrt{5}-1)/2$. The tangent at the point (t^2, t) on the parabola has equation $2ty = x + t^2$, and hence its distance from the origin 0 is $t^2(4t^2+1)^{-1/2}$. Setting this distance equal to 1 produces $X = (b,-\sqrt{b})$ and $Y = (b,\sqrt{b})$ where $b = 2+\sqrt{5}$. Hence $\overrightarrow{OP}\cdot\overrightarrow{OX} = 1$ and so

$$\overrightarrow{OP}\cdot\overrightarrow{PX} = \overrightarrow{OP}\cdot\big(\overrightarrow{OX}-\overrightarrow{OP}\big) = 1-1 = 0.$$

Thus PX is tangent to the circle and, similarly, so also is YQ.

Also solved by 21 other readers. The problem turned out to be rather elementary.

ANSWER TO PHOTO ON PAGE 330

It isn't often that four heads of the same department (at different times) can stand next to each other and smile. From left to right: Heini Halberstam, Paul Bateman, Stewart Cairns, and Mahlon Day. The department is at the University of Illinois; the order of seniority is Cairns, Day, Bateman, and Halberstam (the present head). The picture was taken by Robert J. McEliece.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

A UNIVERSAL ENTIRE FUNCTION

CHARLES BLAIR
Department of Business Administration, University of Illinois, Urbana, IL 61801

LEE A. RUBEL
Department of Mathematics, University of Illinois, Urbana, IL 61801

THEOREM. *There exists an entire function f such that the set $\{f^{(n)}: n = 0, 1, 2, \dots\}$ of all derivatives of f is dense in the space of all entire functions in the topology of uniform convergence on compact subsets of the complex plane $\mathbb{C}$.*

This means that for every compact set $K \subseteq \mathbb{C}$, every $\varepsilon > 0$, and every entire function g, there is a derivative $f^{(N)}$ of f so that $|f^{(N)}(z) - g(z)| < \varepsilon$ for all $z \in K$. We call such an f "universal." A different kind of universality, namely, density of the set of integer *translates*, was treated in **[BIR]** and **[SEW]**. For a treatment of density of *linear combinations* of certain derivatives, see **[RUT]**.

Proof of Theorem. Let $P_1, P_2, P_3, \dots$ be an enumeration of all the polynomials with rational coefficients. Let I be the integral operator $I(h)(z) = \int_0^z h(w)\,dw$. Represent repeated applications of I by I^N. Our function f will have the form

$$f = \sum I^{K_n}(P_n),$$

where the K_n are large positive integers chosen to satisfy the following. First we need $K_n > K_j + \deg P_j$ for $j = 1, 2, \dots, n-1$. Also, let $H_n = I^{K_n}(P_n)$. Then we require that

$$(*) \qquad |H_n^{(j)}(z)| \leqslant \frac{1}{2^n} \quad \text{for} \quad j = 0, 1, 2, \dots, K_{n-1}, |z| \leqslant n.$$

If this can be done, then the series defining f converges uniformly on compact sets and can be differentiated term by term. Further,

$$f^{(K_n)}(z) = P_n(z) + E_n(z)$$

where $|E_n(z)| \leqslant 2^{-(n-1)}$ for $|z| \leqslant n$.

The result follows easily since any entire function g is the uniform limit on compact sets of a sequence of polynomials (the partial sums of its Taylor series) and since on any compact set, any polynomial is uniformly approximated by a polynomial with rational coefficients.

To complete the proof we must show that K_n can be chosen so that (*) is satisfied.

Now $I(z^r) = z^{r+1}/(r+1)$ so that

$$|I^k(z^r)| = \left|\frac{z^{r+k}}{(r+1)\cdots(r+k)}\right| \leqslant |z^r|\frac{|z|^k}{k!}.$$

On any fixed disk $\{|z| \leqslant R\}$, $|z^r| \leqslant R^r$ and $|z^k|/k!$ tends uniformly to zero on this disk as $k \to \infty$. Therefore (*) is easy to achieve by taking K_n sufficiently large, because P_n is just a finite linear combination of terms z^r. The proof is done.

References

[BIR] G. D. Birkhoff, Dèmonstration d'un thèoréme élémentaire sur les fonctions entières, C. R. Acad. Sci.,

Paris, 189(1929) 473–475.
[RUT] L. A. Rubel and B. A. Taylor, A completeness theorem for entire functions, J. Indian Math. Soc., 32(1968) 191–198.
[SEW] W. P. Seidel and J. L. Walsh, On approximation by Euclidean and non-Euclidean translates of an analytic function, Bull. Amer. Math. Soc., 47(1941) 916–920.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

FROM LOSS OF MEMORY TO POISSON

BRUCE R. JOHNSON
Department of Mathematics, University of Victoria, Victoria, B.C., Canada

The purpose of this note is to share my favorite way of presenting the Poisson process and deriving the Poisson distribution for upper division courses in probability or mathematical statistics. Although the ingredients of this approach are well known, they are blended in ways different from standard textbook presentations. The main feature of our approach lies in the formulation of Poisson postulates that have immediate intuitive appeal, avoiding the usual analytic expressions

$$P_1(h) = \lambda h + o(h) \quad \text{and} \quad \sum_{n=2}^{\infty} P_n(h) = o(h).$$

While the principal goal is to derive the Poisson distribution from a foundation of intuitively attractive postulates, our derivation also obtains the exponential and gamma waiting time distributions as important by-products.

Our approach requires the "lack of memory" characterization of the exponential distribution, and the probability distribution of independent and identically distributed (i.i.d.) exponential random variables. The "lack of memory" characterization is an intriguing topic for lecture, but, with appropriate hints, it could be assigned as a challenging homework problem. Finding the distribution of the sum of i.i.d. exponential random variables is an ideal homework problem, either to illustrate the power of generating function theory or to reinforce convolution properties and the tool of mathematical induction.

The presentation begins with an intuitive definition of a Poisson process: a process that continues over time (or space) where certain rare "happenings" occur at random in time (or space) at some fixed average rate. Then several examples are given to develop understanding and to display wide applicability for the Poisson model.

To formalize the definition of a Poisson process, we specify three postulates that are consistent with the intuitive definition.

Poisson Postulates

Stationarity: The probability distribution of the number of happenings in a specified time interval depends only on the length of the interval and not on its position.

Independence: Events involving the number of happenings in disjoint time intervals are independent.

Paris, 189(1929) 473–475.

[RUT] L. A. Rubel and B. A. Taylor, A completeness theorem for entire functions, J. Indian Math. Soc., 32(1968) 191–198.

[SEW] W. P. Seidel and J. L. Walsh, On approximation by Euclidean and non-Euclidean translates of an analytic function, Bull. Amer. Math. Soc., 47(1941) 916–920.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

FROM LOSS OF MEMORY TO POISSON

BRUCE R. JOHNSON
Department of Mathematics, University of Victoria, Victoria, B.C., Canada

The purpose of this note is to share my favorite way of presenting the Poisson process and deriving the Poisson distribution for upper division courses in probability or mathematical statistics. Although the ingredients of this approach are well known, they are blended in ways different from standard textbook presentations. The main feature of our approach lies in the formulation of Poisson postulates that have immediate intuitive appeal, avoiding the usual analytic expressions

$$P_1(h) = \lambda h + o(h) \quad \text{and} \quad \sum_{n=2}^{\infty} P_n(h) = o(h).$$

While the principal goal is to derive the Poisson distribution from a foundation of intuitively attractive postulates, our derivation also obtains the exponential and gamma waiting time distributions as important by-products.

Our approach requires the "lack of memory" characterization of the exponential distribution, and the probability distribution of independent and identically distributed (i.i.d.) exponential random variables. The "lack of memory" characterization is an intriguing topic for lecture, but, with appropriate hints, it could be assigned as a challenging homework problem. Finding the distribution of the sum of i.i.d. exponential random variables is an ideal homework problem, either to illustrate the power of generating function theory or to reinforce convolution properties and the tool of mathematical induction.

The presentation begins with an intuitive definition of a Poisson process: a process that continues over time (or space) where certain rare "happenings" occur at random in time (or space) at some fixed average rate. Then several examples are given to develop understanding and to display wide applicability for the Poisson model.

To formalize the definition of a Poisson process, we specify three postulates that are consistent with the intuitive definition.

Poisson Postulates

Stationarity: The probability distribution of the number of happenings in a specified time interval depends only on the length of the interval and not on its position.

Independence: Events involving the number of happenings in disjoint time intervals are independent.

Rareness: Beginning at any starting time there is certain to be a positive (but finite) wait until the next happening, and multiple happenings cannot occur simultaneously.

Now, a Poisson process is defined formally to be any process that continues over time (or space) and satisfies the Poisson postulates.

For a given Poisson process we reset the time clock to zero, let

$$T = \text{waiting time until the next happening},$$

and investigate the probability distribution of T. By the rareness postulate, T is seen to be a positive valued random variable. The tail probabilities $P(T > t)$ are positive for every $t \geqslant 0$. This follows from the fact that

$$P(T > s) \uparrow P(T > 0) = 1 \quad \text{as} \quad s \downarrow 0,$$

so we can find $s_0 > 0$ such that $P(T > s_0) > 0$, from which we use independence and stationarity to conclude

$$\begin{aligned} P(T > ns_0) &= P\big(\text{no happenings in } (0, ns_0]\big) \\ &= \big(P\big(\text{no happenings in } (0, s_0]\big)\big)^n > 0 \end{aligned}$$

for all positive integers n. Also, T has the "lack of memory" property because, for $s \geqslant 0$ and $t \geqslant 0$, we obtain

$$\begin{aligned} P(T > s + t | T > t) &= \frac{P\big(\text{no happenings in } (0, s + t]\big)}{P(T > t)} \\ &= \frac{P\big(\text{no happenings in } (0, t]\big)P\big(\text{no happenings in } (0, s]\big)}{P(T > t)} \\ &\quad \text{by independence and stationarity,} \\ &= P(T > s). \end{aligned}$$

Therefore, T must have the exponential distribution with parameter $\lambda = -\ln P(T > 1)$.

Letting $S_n = T_1 + T_2 + \cdots + T_n$ where $T_1 = T$ and T_j denotes the waiting time between the $(j - 1)$th and the jth happening, for $2 \leqslant j \leqslant n$, we use stationarity and independence to conclude that S_n is the sum of n independent random variables each having the exponential (λ) distribution. Hence, the distribution of S_n is given by the gamma density

$$f_n(x) = \frac{\lambda^n}{(n - 1)!} x^{n-1} e^{-\lambda x}, \quad x > 0.$$

If, for $t > 0$, we let

$$N(t) = \text{the number of happenings in time interval } (0, t],$$

then

$$P(N(t) = 0) = P(T > t) = e^{-\lambda t},$$

and, for $n = 1, 2, 3, \ldots,$

$$\begin{aligned} P(N(t) = n) &= P(S_n \leqslant t, \quad S_{n+1} > t) \\ &= P(S_n \leqslant t) - P(S_{n+1} \leqslant t) \\ &= \int_0^t \frac{\lambda^n}{(n - 1)!} x^{n-1} e^{-\lambda x}\, dx - \int_0^t \frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x}\, dx \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \end{aligned}$$

from integration by parts applied to the first integral. Thus, whenever the Poisson postulates are satisfied, $N(t)$ has the Poisson distribution with mean $\mu = \lambda t$. Since $E(N(1)) = \lambda$, we see that λ is just the average number of happenings per unit time (or space).

References

1. K. L. Chung, Elementary Probability Theory with Stochastic Processes, 3rd ed., Springer-Verlag, New York, 1979.
2. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., Wiley, New York, 1968.
3. D. A. S. Fraser, Probability and Statistics Theory and Applications, Duxbury, Belmont, 1976.
4. B. W. Lindgren, Statistical Theory, 3rd ed., Macmillan, New York, 1976.
5. E. Parzen, Modern Probability Theory and Its Applications, Wiley, New York, 1960.
6. S. M. Ross, Introduction to Probability Models, 2nd ed., Academic Press, New York, 1980.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2995 *Proposed by Miroslav D. Ašić, London School of Economics.*

A square matrix of order n, $n \geqslant 2$, is said to be "good" if it is symmetric, invertible and all its entries are positive. What is the largest possible number of zero entries in the inverse of a "good" matrix?

E 2996. *Proposed by Phil Novinger and Daniel Oberlin, Florida State University.*

Let $\sum_{n=1}^{\infty} a_n$ be a positive term convergent series and $r_n = \sum_{m=n}^{\infty} a_m$, $n = 1, 2, \ldots$.

(i) Show that if $0 < p < 1$, then there is an absolute constant C_p such that

$$\sum_{n=1}^{\infty} \frac{a_n}{r_n^p} < C_p \left[\sum_{n=1}^{\infty} a_n \right]^{1-p}$$

from integration by parts applied to the first integral. Thus, whenever the Poisson postulates are satisfied, $N(t)$ has the Poisson distribution with mean $\mu = \lambda t$. Since $E(N(1)) = \lambda$, we see that λ is just the average number of happenings per unit time (or space).

References

1. K. L. Chung, Elementary Probability Theory with Stochastic Processes, 3rd ed., Springer-Verlag, New York, 1979.
2. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., Wiley, New York, 1968.
3. D. A. S. Fraser, Probability and Statistics Theory and Applications, Duxbury, Belmont, 1976.
4. B. W. Lindgren, Statistical Theory, 3rd ed., Macmillan, New York, 1976.
5. E. Parzen, Modern Probability Theory and Its Applications, Wiley, New York, 1960.
6. S. M. Ross, Introduction to Probability Models, 2nd ed., Academic Press, New York, 1980.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2995 *Proposed by Miroslav D. Ašić, London School of Economics.*

A square matrix of order n, $n \geqslant 2$, is said to be "good" if it is symmetric, invertible and all its entries are positive. What is the largest possible number of zero entries in the inverse of a "good" matrix?

E 2996. *Proposed by Phil Novinger and Daniel Oberlin, Florida State University.*

Let $\sum_{n=1}^{\infty} a_n$ be a positive term convergent series and $r_n = \sum_{m=n}^{\infty} a_m$, $n = 1, 2, \ldots$.

(i) Show that if $0 < p < 1$, then there is an absolute constant C_p such that

$$\sum_{n=1}^{\infty} \frac{a_n}{r_n^p} < C_p \left[\sum_{n=1}^{\infty} a_n \right]^{1-p}$$

for all such series.

(ii) Find the best possible C_p.

E 2997. *Proposed by Irving Adler, North Bennington, Vermont.*

Let p_0 be the perimeter of an inscribed regular n-gon in a unit circle, and let d_k be the distance from the center of the circle to the side of the inscribed regular $(2^k \cdot n)$-gon. Prove that

$$\frac{p_0}{2}\prod_{k=1}^{\infty}\frac{1}{d_k}=\pi.$$

E 2998. *Proposed by Clark Kimberling, University of Evansville.*

Suppose x and y are complex numbers such that $(x^n - y^n)/(x-y)$ is an integer for some four consecutive positive integers n. Prove it is an integer for all positive integers n.

E 2999. *Proposed by J. D. Shallit, University of California, Berkeley, and Karel Zikan, San Jose State University.*

Let S be the set of nontrivial integer kth powers, i.e.,

$$S=\{n^k \mid n\geqslant 2, k\geqslant 2\}=\{4,8,9,16,25,27,32,36,\dots\}.$$

Evaluate $\Sigma(s-1)^{-1}$, the sum being extended over all members s of S.

E 3000. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

Let $0<u_1<u_2<\cdots$ be an infinite sequence; suppose $\Sigma 1/u_i$ converges. Denote by $f(x)$ the number of pairs (i,j) for which the partial sum $\Sigma_i^j u_r$ is $\leqslant x$. Prove that $\lim f(x)/x=0$ $(x\to\infty)$.

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Permutations with f Fixed Points

S 14* [1979, 503]. *Proposed by C. L. Mallows, Bell Laboratories, Murray Hill, NJ.*

Let $n(m,f,r)$ be the number of arrangements $(a_1,a_2,\dots,a_m)$ of $(1,2,\dots,m)$ that have f fixed points $(a_i=i)$ and r *rises* $(a_i<a_{i+1})$. Prove (a) $n(m,0,r)=n(m,1,r)$ for $1\leqslant r\leqslant m-1$, and (b) $n(m,f,r)=$

$$\sum_{j=2}^{m-r}(-1)^{m-r-j}(j-1)^j j^{m-f-j}\binom{f+j-1}{j-1}\binom{m+1}{m-r-j}+(-1)^{m-f}\frac{\delta_{r+1-f}}{m-f+1}\binom{m}{f}$$

for $0\leqslant f\leqslant r+1\leqslant m, 0\leqslant r$, where $\delta_k=1,=0$ if $k=0,\neq 0$, respectively.

This problem has been solved by *Ira Gessel, Massachusetts Institute of Technology*. Part (a) can be generalized as follows: Let A_f be the set of arrangements of $(1,2,\dots,m)$ with f fixed points. For any subset R of $\{1,2,\dots,m-1\}$, let B_R be the set of arrangements with $\{i \mid a_i<a_{i+1}\}=R$. Then if R is nonempty $|A_0\cap B_R|=|A_1\cap B_R|$. Part (b) can be generalized to the formula

$$n(m,f,r)=\sum_{j=2}^{m-r}(-1)^{m-r-j}\left[(j-1)^j-E_j\right]j^{m-f-j}\binom{f+j-1}{j-1}\binom{m+1}{m-r-j},$$

where $E_j=\Sigma_{k=m-f+1}^{j}(-1)^k j^{j-k}\binom{j}{k}$, for $0\leqslant r<m-1$.

The solution will appear elsewhere.

for all such series.

(ii) Find the best possible C_p.

E 2997. *Proposed by Irving Adler, North Bennington, Vermont.*

Let p_0 be the perimeter of an inscribed regular n-gon in a unit circle, and let d_k be the distance from the center of the circle to the side of the inscribed regular $(2^k \cdot n)$-gon. Prove that

$$\frac{p_0}{2}\prod_{k=1}^{\infty}\frac{1}{d_k} = \pi.$$

E 2998. *Proposed by Clark Kimberling, University of Evansville.*

Suppose x and y are complex numbers such that $(x^n - y^n)/(x - y)$ is an integer for some four consecutive positive integers n. Prove it is an integer for all positive integers n.

E 2999. *Proposed by J. D. Shallit, University of California, Berkeley, and Karel Zikan, San Jose State University.*

Let S be the set of nontrivial integer kth powers, i.e.,

$$S = \{n^k | n \geqslant 2, k \geqslant 2\} = \{4,8,9,16,25,27,32,36,\dots\}.$$

Evaluate $\Sigma(s-1)^{-1}$, the sum being extended over all members s of S.

E 3000. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

Let $0 < u_1 < u_2 < \cdots$ be an infinite sequence; suppose $\Sigma 1/u_i$ converges. Denote by $f(x)$ the number of pairs (i, j) for which the partial sum $\Sigma_i^j u_r$ is $\leqslant x$. Prove that $\lim f(x)/x = 0$ $(x \to \infty)$.

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Permutations with f Fixed Points

S 14* [1979, 503]. *Proposed by C. L. Mallows, Bell Laboratories, Murray Hill, NJ.*

Let $n(m,f,r)$ be the number of arrangements $(a_1, a_2,\dots, a_m)$ of $(1,2,\dots, m)$ that have f fixed points $(a_i = i)$ and r *rises* $(a_i < a_{i+1})$. Prove (a) $n(m,0,r) = n(m,1,r)$ for $1 \leqslant r \leqslant m-1$, and (b) $n(m,f,r) =$

$$\sum_{j=2}^{m-r}(-1)^{m-r-j}(j-1)^j j^{m-f-j}\binom{f+j-1}{j-1}\binom{m+1}{m-r-j} + (-1)^{m-f}\frac{\delta_{r+1-f}}{m-f+1}\binom{m}{f}$$

for $0 \leqslant f \leqslant r+1 \leqslant m, 0 \leqslant r$, where $\delta_k = 1, = 0$ if $k = 0, \neq 0$, respectively.

This problem has been solved by *Ira Gessel, Massachusetts Institute of Technology*. Part (a) can be generalized as follows: Let A_f be the set of arrangements of $(1,2,\dots, m)$ with f fixed points. For any subset R of $\{1,2,\dots, m-1\}$, let B_R be the set of arrangements with $\{i | a_i < a_{i+1}\} = R$. Then if R is nonempty $|A_0 \cap B_R| = |A_1 \cap B_R|$. Part (b) can be generalized to the formula

$$n(m,f,r) = \sum_{j=2}^{m-r}(-1)^{m-r-j}\left[(j-1)^j - E_j\right]j^{m-f-j}\binom{f+j-1}{j-1}\binom{m+1}{m-r-j},$$

where $E_j = \Sigma_{k=m-f+1}^{j}(-1)^k j^{j-k}\binom{j}{k}$, for $0 \leqslant r < m-1$.

The solution will appear elsewhere.

SOLUTIONS OF ELEMENTARY PROBLEMS

Series Σa_n and Σna_n

E 2902 [1981, 619]. *Proposed by Lynn Sennott, Illinois State University.*

Let $\Sigma a_n = a$ be a convergent series of nonnegative terms. Let s_n be the nth partial sum. Show that Σna_n converges if and only if $\Sigma(a - s_n)$ converges.

Solution. Many solvers pointed out that the proposal is a version of Fubini's theorem. Some gave generalizations to Lebesgue-Stieltjes integrals; some replaced Σna_n by $\Sigma n^k a_n$, with appropriate changes in partial sums. One of the more extensive generalizations was submitted by *J.R. Nurcombe, Hall Green Technical College, Birmingham, England.*

THEOREM. *Let Σa_k be a convergent series of real numbers.* (See remarks at end.) *Set $r_n = \Sigma_n^\infty a_k$. Suppose one of* (i), (ii), (iii) *holds.*

(i) *$\{Q_k\}$ is a bounded sequence of real numbers.*

(iia) *$\Sigma_1^\infty |Q_{n+k}^{-1} - Q_{n+k-1}^{-1}| = O(|Q_n^{-1}|)$, and*

(iib) *$\Sigma_0^n |r_{k+1}^{-1} - r_k^{-1}| = O(|r_{n+1}^{-1}|)$, as $n \to \infty$.*

(iii) *$\{Q_{k-1}a_k\}$ are all positive; $\{|Q_k|\}$ is nondecreasing; and* (iib) *holds.*

Define q_k by the relation $Q_n = \Sigma_0^n q_k$.

Conclusion: $\Sigma_1^\infty Q_{k-1}a_k$ and $\Sigma_0^\infty q_k r_{k+1}$ converge to the same limit.

We use a well-known lemma (K. Knopp, *Theory and application of infinite series*, Hafner, New York, 1971, p. 130). If $\{p_k\}$ is a sequence of real numbers with $|p_n| \to \infty$ $(n \to \infty)$, and if $\Sigma_0^n |p_k - p_{k-1}| = O(|p_n|)$, and if Σa_k converges, then $\Sigma_0^n p_k a_k = o(p_n)$. We also need a lemma that seems to be new.

LEMMA. *If $\{m_k\}$ is a real sequence, if $M_n = \Sigma_0^n m_k$, and if $\Sigma_1^\infty |m_{n+k}| = O(|M_n|)$ $(n \to \infty)$, and if Σa_k converges, then $\Sigma_n^\infty M_k a_k = o(M_n)$.*

Proof of the lemma. We have

$$\sum_n^\infty M_k a_k = \sum_n^\infty M_k(r_k - r_{k+1}) = M_n r_n + \sum_{k=1}^\infty r_{n+k}(M_{n+k} - M_{n+k-1}) = M_n r_n + \sum_{k=1}^\infty m_{n+k} r_{n+k}.$$

Hence

$$\sum_{k=n}^\infty M_k a_k / M_n = r_n + \sum_{k=1}^\infty m_{n+k} r_{n+k} / M_n.$$

Since Σa_k converges, we know that $r_n \to 0$ $(n \to \infty)$. Also,

$$\left| \sum_{k=1}^\infty m_{n+k} r_{n+k} / M_n \right| \leqslant \sum |m_{n+k} r_{n+k} / M_n| \leqslant \left(\max_{k \geqslant 1} |r_{n+k}| \right) \cdot \left(\sum |m_{n+k}| / |M_n| \right).$$

The second factor here is bounded by hypothesis, and the first factor $\to 0$ for every k (as $n \to \infty$). The lemma is proved.

Proof of the theorem. For all three parts (i), (ii), (iii), we need the identity

$$(*) \qquad \sum_1^n Q_{k-1} a_k = \sum_0^n q_k r_{k+1} - Q_n r_{n+1}.$$

(i) Note $r_n \to 0$. Since $\{Q_k\}$ is bounded, $Q_n r_{n+1} \to 0$. Thus $(*)$ shows that if $\Sigma Q_{k-1} a_k$ converges, then so does $\Sigma q_k r_{k+1}$ (and conversely), and to the same sum.

(ii) Suppose $\sum q_k r_{k+1}$ converges to L. In the "well-known" lemma above, set $p_k = 1/r_{k+1}$; then

$$\sum_0^n q_k r_{k+1}/r_{k+1} = o(1/r_{n+1}).$$

This is tantamount to $Q_n r_{n+1} \to 0$. From $(*)$, $\sum Q_{k-1} a_k$ converges to L. For the converse, if $\sum Q_{k-1} a_k$ converges, take $M_k = 1/Q_{k-1}$ in the second lemma. Then

$$\sum_{n+1}^{\infty} Q_{k-1} a_k / Q_{k-1} = o(1/Q_n),$$

so that $Q_n r_{n+1} \to 0$. The conclusion follows from $(*)$.

(iii) If $\sum q_k r_{k+1}$ converges, we repeat the pertinent (first) half of (ii). For the converse, if $\sum Q_{k-1} a_k$ converges, the inequalities

$$|Q_n r_{n+1}| = |Q_n| \left| \sum_{n+1}^{\infty} a_k \right| \leqslant |Q_n| \sum_{n+1}^{\infty} |a_k| \leqslant \sum_{n+1}^{\infty} |Q_{k-1}||a_k| = \sum_{n+1}^{\infty} Q_{k-1} a_k$$

hold, because $\{|Q_k|\}$ is nondecreasing and $\{Q_{k-1} a_k\}$ is positive. Thus $Q_n r_{n+1} \to 0$. The identity $(*)$ completes the proof. □

Remarks. The result of proposal E 2902 follows from (ii) or (iii) if $q_n \equiv 1$.

The condition $\Sigma_0^n |r_{k+1}{}^{-1} - r_k{}^{-1}| = O(|r_{n+1}{}^{-1}|)$ allows the sequence $\{a_k\}$ to be of mixed sign. For example, any convergent geometric series with negative ratio does satisfy the condition.

The conditions of the two lemmas are necessary as well as sufficient for the respective conclusions.

With minor changes, the theorem continues to hold for a sequence of complex numbers.

Also solved by 117 readers.

Recoverable Primes

E 2904 [1981, 619]. *Proposed by D. Hensley, Texas A&M University.*

Let $p > 2$ be prime. Call x *recoverable* if $x^{2^v} \equiv x \bmod p$ for some $v \geqslant 1$. Show that the set of recoverable x is permuted by squaring.

Solution by D. M. Bloom, Brooklyn College. The case $x \equiv 0$ is easy: $0^2 = 0$. For $x \not\equiv 0$, "x recoverable" means that the residue class of x (mod p) has odd order in the group Z_p^* of nonzero residue classes: $x^{2^v - 1} = 1$.

LEMMA. *The elements of odd order in any group are permuted by squaring* (f: $x \to x^2$).

Proof. If x has order $2m+1$, then (a) x^2 has order $2m+1$, (b) x^{m+1} has order $2m+1$ and its square is x, and (c) if $y^2 = x$ where y has odd order, then necessarily $y^{2m+1} = 1$ and hence $y = y^{2(m+1)} = x^{m+1}$. Statements (a), (b) and (c) establish the assertion.

Brandler noted that if G is a group and if $(m, |G|) = 1$, then f: $G \to G$ ($f(g) \equiv g^m$) is bijective. Cater and Erlebach noted that the assertion is true in any field.

Also solved by R. Beigel (student), J. Brandler, J. Browkin (Poland), K. A. Brown, P. S. Bruckman, P. Butzmann (Germany), F. S. Cater, Chico Problem Group, G. Dedaj (student), J. Deutsch, C. W. Dodge, H. M. Edgar, L. Erlebach, P. Flusser, L. L. Foster, C. V. Heuer, J. Hook (student), Y. Ikeda, W. Janous (Austria), M. Josephy (Costa Rica), O. P. Lossers (Netherlands), J. M. Masley, D. E. Orr, J. P. Robertson, J. Shallit, D. Singmaster (U.K.), L. Somer, K. Spindler (Germany), A. Stenger, G. Stutzin (student), M. B. Suryanarayana, R. Teitler (U.K.), E. Trost (Switzerland), D. G. Weinman, M. Woltermann, K. Zikan, and the proposer.

A Sequence of Angles in a Triangle Converging to $\pi/3$

E 2906 [1981, 620]. *Proposed by Jack Garfunkel, Flushing, NY.*

Let I be the incenter of triangle ABC. A', B', C' are the intersections of AI, BI, CI with the incircle of ABC. Continue the process by defining I' (incenter of $A'B'C'$), then $A''B''C''$, etc. Prove that the angles of triangle $A^{(n)}B^{(n)}C^{(n)}$ approach $\pi/3$.

Solution by Clayton W. Dodge, University of Maine at Orono. From triangle IAC we have that

$$\measuredangle AIC = \pi - A/2 - C/2 = (\pi + B)/2,$$

so that $B' = \measuredangle A'B'C' = \frac{1}{2}\measuredangle AIC = (\pi + B)/4$. Similar relations hold for A' and C'. Assuming $A \leqslant B \leqslant C$, then

$$A' = \tfrac{1}{4}(\pi + A) \leqslant B' \leqslant \tfrac{1}{4}(\pi + B) \leqslant C' = \tfrac{1}{4}(\pi + C), \quad \text{and} \quad C' - A' = (C - A)/4,$$

so triangle $A'B'C'$ is "4 times closer" to equilateral than triangle ABC is. The theorem follows.

Also solved by A. Bager (Denmark), N. M. Bejlegaard (Denmark), P. S. Bruckman, G. Dedaj (student), J. Dou (Spain), J. Heuver (Canada), W. Janous (Austria), P. Kumar (India), O. P. Lossers (Netherlands), E. Nadler, W. A. Newcomb, R. E. Spaulding, J. M. Stark, D. Tsakmakis (Greece), G. H. Wenzel (Germany), and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1983. The solver's full post-office address should be on each sheet.

6427. *Proposed by Fred Galvin, University of Kansas.*

Suppose that S is a semigroup, and there are elements a and b in S such that (1) for each x in S there is a unique y in S such that $xy = a$, and (2) for each x in S there is a unique y in S such that $yx = b$. Show that S is a group.

6428. *Proposed by M. Steele, University of Chicago.*

Let ε be a separation of $\{1,2,\ldots,2n\}$ into two disjoint classes $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots < b_n$. Let M_k denote the number of solutions of $a_i - b_j = k$, $-2n < k < 2n$, and put $M(n) = \min_\varepsilon \max_k M_k$. Selfridge, Motzkin, and Ralston showed $\liminf n^{-1}M(n) \leqslant .4$, and Moser showed $\limsup n^{-1}M(n) \geqslant .357$. Prove that the limit $\lim n^{-1}M(n)$ actually exists.

6429. *Proposed by B. Powell, Kirkland, WA.*

Prove that for any positive integer M there exists an even positive integer $n_0 = n_0(M)$ such that the number of pairs of consecutive primes (P_j, P_{j+1}) that differ by exactly n_0 is greater than M.

SOLUTIONS OF ADVANCED PROBLEMS

A Trace Inequality

6360 [1981, 623]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that

$$\operatorname{tr}\left[(rI - A)^T \operatorname{adj}(rI - A)\right] \leqslant 0,$$

where I is the $n \times n$ identity matrix, adj X is the classical adjoint of X, X^T is the transpose of X, and tr X is the trace of X. Find necessary and sufficient conditions for equality to hold.

A Sequence of Angles in a Triangle Converging to $\pi/3$

E 2906 [1981, 620]. *Proposed by Jack Garfunkel, Flushing, NY.*

Let I be the incenter of triangle ABC. A', B', C' are the intersections of AI, BI, CI with the incircle of ABC. Continue the process by defining I' (incenter of $A'B'C'$), then $A''B''C''$, etc. Prove that the angles of triangle $A^{(n)}B^{(n)}C^{(n)}$ approach $\pi/3$.

Solution by Clayton W. Dodge, University of Maine at Orono. From triangle IAC we have that

$$\measuredangle AIC = \pi - A/2 - C/2 = (\pi + B)/2,$$

so that $B' = \measuredangle A'B'C' = \frac{1}{2}\measuredangle AIC = (\pi + B)/4$. Similar relations hold for A' and C'. Assuming $A \leqslant B \leqslant C$, then

$$A' = \tfrac{1}{4}(\pi + A) \leqslant B' \leqslant \tfrac{1}{4}(\pi + B) \leqslant C' = \tfrac{1}{4}(\pi + C), \quad \text{and} \quad C' - A' = (C - A)/4,$$

so triangle $A'B'C'$ is "4 times closer" to equilateral than triangle ABC is. The theorem follows.

Also solved by A. Bager (Denmark), N. M. Bejlegaard (Denmark), P. S. Bruckman, G. Dedaj (student), J. Dou (Spain), J. Heuver (Canada), W. Janous (Austria), P. Kumar (India), O. P. Lossers (Netherlands), E. Nadler, W. A. Newcomb, R. E. Spaulding, J. M. Stark, D. Tsakmakis (Greece), G. H. Wenzel (Germany), and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1983. The solver's full post-office address should be on each sheet.

6427. *Proposed by Fred Galvin, University of Kansas.*

Suppose that S is a semigroup, and there are elements a and b in S such that (1) for each x in S there is a unique y in S such that $xy = a$, and (2) for each x in S there is a unique y in S such that $yx = b$. Show that S is a group.

6428. *Proposed by M. Steele, University of Chicago.*

Let ε be a separation of $\{1,2,\ldots,2n\}$ into two disjoint classes $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots < b_n$. Let M_k denote the number of solutions of $a_i - b_j = k$, $-2n < k < 2n$, and put $M(n) = \min_\varepsilon \max_k M_k$. Selfridge, Motzkin, and Ralston showed $\liminf n^{-1}M(n) \leqslant .4$, and Moser showed $\limsup n^{-1}M(n) \geqslant .357$. Prove that the limit $\lim n^{-1}M(n)$ actually exists.

6429. *Proposed by B. Powell, Kirkland, WA.*

Prove that for any positive integer M there exists an even positive integer $n_0 = n_0(M)$ such that the number of pairs of consecutive primes (P_j, P_{j+1}) that differ by exactly n_0 is greater than M.

SOLUTIONS OF ADVANCED PROBLEMS

A Trace Inequality

6360 [1981, 623]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that

$$\operatorname{tr}\left[(rI - A)^T \operatorname{adj}(rI - A)\right] \leqslant 0,$$

where I is the $n \times n$ identity matrix, adj X is the classical adjoint of X, X^T is the transpose of X, and tr X is the trace of X. Find necessary and sufficient conditions for equality to hold.

A Sequence of Angles in a Triangle Converging to $\pi/3$

E 2906 [1981, 620]. *Proposed by Jack Garfunkel, Flushing, NY.*

Let I be the incenter of triangle ABC. A', B', C' are the intersections of AI, BI, CI with the incircle of ABC. Continue the process by defining I' (incenter of $A'B'C'$), then $A''B''C''$, etc. Prove that the angles of triangle $A^{(n)}B^{(n)}C^{(n)}$ approach $\pi/3$.

Solution by Clayton W. Dodge, University of Maine at Orono. From triangle IAC we have that

$$\measuredangle AIC = \pi - A/2 - C/2 = (\pi + B)/2,$$

so that $B' = \measuredangle A'B'C' = \frac{1}{2}\measuredangle AIC = (\pi + B)/4$. Similar relations hold for A' and C'. Assuming $A \leqslant B \leqslant C$, then

$$A' = \tfrac{1}{4}(\pi + A) \leqslant B' \leqslant \tfrac{1}{4}(\pi + B) \leqslant C' = \tfrac{1}{4}(\pi + C), \quad \text{and} \quad C' - A' = (C - A)/4,$$

so triangle $A'B'C'$ is "4 times closer" to equilateral than triangle ABC is. The theorem follows.

Also solved by A. Bager (Denmark), N. M. Bejlegaard (Denmark), P. S. Bruckman, G. Dedaj (student), J. Dou (Spain), J. Heuver (Canada), W. Janous (Austria), P. Kumar (India), O. P. Lossers (Netherlands), E. Nadler, W. A. Newcomb, R. E. Spaulding, J. M. Stark, D. Tsakmakis (Greece), G. H. Wenzel (Germany), and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1983. The solver's full post-office address should be on each sheet.

6427. *Proposed by Fred Galvin, University of Kansas.*

Suppose that S is a semigroup, and there are elements a and b in S such that (1) for each x in S there is a unique y in S such that $xy = a$, and (2) for each x in S there is a unique y in S such that $yx = b$. Show that S is a group.

6428. *Proposed by M. Steele, University of Chicago.*

Let ε be a separation of $\{1,2,\ldots,2n\}$ into two disjoint classes $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots < b_n$. Let M_k denote the number of solutions of $a_i - b_j = k$, $-2n < k < 2n$, and put $M(n) = \min_\varepsilon \max_k M_k$. Selfridge, Motzkin, and Ralston showed $\liminf n^{-1}M(n) \leqslant .4$, and Moser showed $\limsup n^{-1}M(n) \geqslant .357$. Prove that the limit $\lim n^{-1}M(n)$ actually exists.

6429. *Proposed by B. Powell, Kirkland, WA.*

Prove that for any positive integer M there exists an even positive integer $n_0 = n_0(M)$ such that the number of pairs of consecutive primes (P_j, P_{j+1}) that differ by exactly n_0 is greater than M.

SOLUTIONS OF ADVANCED PROBLEMS

A Trace Inequality

6360 [1981, 623]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that

$$\operatorname{tr}\left[(rI - A)^T \operatorname{adj}(rI - A)\right] \leqslant 0,$$

where I is the $n \times n$ identity matrix, adj X is the classical adjoint of X, X^T is the transpose of X, and tr X is the trace of X. Find necessary and sufficient conditions for equality to hold.

Solution by the proposer. Let R^n denote the vector space of real n-tuples. For $y \in R^n$, we denote by y_i the ith component of y and by $y > 0$ we mean that $y_i > 0$ for $i = 1, 2, \ldots, n$.

Friedland and Karlin have proved [Duke Math. J., 42 (1975) 459–490, Theorem 3.1] that if A is a nonnegative irreducible $n \times n$ matrix with Perron root r and if

$$Au = ru, Av = rv, u, v \in R^n, u > 0, v > 0, u^Tv = 1,$$

then

$$r \leqslant \prod_{i=1}^{n} \left(\frac{(Ax)_i}{x_i} \right)^{u_i v_i} \tag{1}$$

whenever $x \in R^n$, $x > 0$. Equality holds in (1) if and only if x and u are linearly dependent. By the arithmetic-geometric mean inequality, we have

$$\prod_{i=1}^{n} \left(\frac{(Ax)_i}{x_i} \right)^{u_i v_i} \leqslant \sum_{i=1}^{n} u_i v_i \frac{(Ax)_i}{x_i}$$

with equality if and only if all the $(Ax)_i/x_i$ are equal, i.e., if and only if x is a positive eigenvector of A, i.e., if and only if x and u are linearly dependent. Consequently, a corollary of (1) is

$$r \leqslant \sum_{i=1}^{n} u_i v_i \frac{(Ax)_i}{x_i} \quad \text{for all } x > 0 \tag{2}$$

with equality if and only if x and u are linearly dependent. Taking $x = v$ in (2), we obtain

$$r \leqslant \sum_{i=1}^{n} u_i (Av)_i = u^T A v \tag{3}$$

with equality if and only if u and v are linearly dependent.

It is known that

$$B := \operatorname{adj}(rI - A) = \beta u v^T \quad \text{for some } \beta > 0.$$

Then, by (3), we have

$$\begin{aligned} \operatorname{tr}\left[(rI - A)^T B\right] &= \beta \operatorname{tr}\left[(rI - A)^T u v^T\right] \\ &= \beta\left((rI - A)^T u\right)^T v = \beta u^T (rI - A) v = \beta (r - u^T A v) \leqslant 0. \end{aligned}$$

Here we have made use of the fact that the rank one matrix ab^T has trace equal to a^Tb.

Equality holds if and only if u and v are linearly dependent, i.e., if and only if A and A^T have a common eigenvector corresponding to r. This in turn holds if and only if $B := \operatorname{adj}(rI - A)$ is symmetric. It is easy to see that this condition is sufficient, since when it holds

$$(rI - A)^T B = [B(rI - A)]^T = 0.$$

In particular, equality holds whenever A is normal or generalized doubly stochastic.

Bounds of Analytic Functions

6362 [1981, 710]. *Proposed by F. S. Cater, Portland State University.*

Let $F(z)$ be an analytic function for $|z| < 1$. For each number u, $0 \leqslant u < 2\pi$, let

$$h(u) = \sup_{0 < t < 1} |F(te^{iu})| \text{ and } g(u) = \limsup_{t \to 1-} |F(te^{iu})|.$$

Suppose that for each positive integer n, $\int_0^{2\pi} h(u)^n du < \infty$ in the Lebesgue sense. (This would happen, for example, if F were bounded.) Prove that

Solution by the proposer. Let R^n denote the vector space of real n-tuples. For $y \in R^n$, we denote by y_i the ith component of y and by $y > 0$ we mean that $y_i > 0$ for $i = 1,2,\ldots, n$.

Friedland and Karlin have proved [Duke Math. J., 42 (1975) 459–490, Theorem 3.1] that if A is a nonnegative irreducible $n \times n$ matrix with Perron root r and if

$$Au = ru, Av = rv, u, v \in R^n, u > 0, v > 0, u^Tv = 1,$$

then

$$(1) \qquad r \leqslant \prod_{i=1}^{n} \left(\frac{(Ax)_i}{x_i} \right)^{u_i v_i}$$

whenever $x \in R^n$, $x > 0$. Equality holds in (1) if and only if x and u are linearly dependent. By the arithmetic-geometric mean inequality, we have

$$\prod_{i=1}^{n} \left(\frac{(Ax)_i}{x_i} \right)^{u_i v_i} \leqslant \sum_{i=1}^{n} u_i v_i \frac{(Ax)_i}{x_i}$$

with equality if and only if all the $(Ax)_i/x_i$ are equal, i.e., if and only if x is a positive eigenvector of A, i.e., if and only if x and u are linearly dependent. Consequently, a corollary of (1) is

$$(2) \qquad r \leqslant \sum_{i=1}^{n} u_i v_i \frac{(Ax)_i}{x_i} \quad \text{for all } x > 0$$

with equality if and only if x and u are linearly dependent. Taking $x = v$ in (2), we obtain

$$(3) \qquad r \leqslant \sum_{i=1}^{n} u_i (Av)_i = u^T A v$$

with equality if and only if u and v are linearly dependent.

It is known that

$$B := \operatorname{adj}(rI - A) = \beta u v^T \quad \text{for some } \beta > 0.$$

Then, by (3), we have

$$\begin{aligned} \operatorname{tr}\left[(rI - A)^T B\right] &= \beta \operatorname{tr}\left[(rI - A)^T u v^T\right] \\ &= \beta\left((rI - A)^T u\right)^T v = \beta u^T (rI - A) v = \beta(r - u^T A v) \leqslant 0. \end{aligned}$$

Here we have made use of the fact that the rank one matrix ab^T has trace equal to $a^T b$.

Equality holds if and only if u and v are linearly dependent, i.e., if and only if A and A^T have a common eigenvector corresponding to r. This in turn holds if and only if $B := \operatorname{adj}(rI - A)$ is symmetric. It is easy to see that this condition is sufficient, since when it holds

$$(rI - A)^T B = [B(rI - A)]^T = 0.$$

In particular, equality holds whenever A is normal or generalized doubly stochastic.

Bounds of Analytic Functions

6362 [1981, 710]. *Proposed by F. S. Cater, Portland State University.*

Let $F(z)$ be an analytic function for $|z| < 1$. For each number u, $0 \leqslant u < 2\pi$, let

$$h(u) = \sup_{0 < t < 1} |F(te^{iu})| \text{ and } g(u) = \limsup_{t \to 1-} |F(te^{iu})|.$$

Suppose that for each positive integer n, $\int_0^{2\pi} h(u)^n du < \infty$ in the Lebesgue sense. (This would happen, for example, if F were bounded.) Prove that

$$\sup_{0\leqslant u<2\pi} g(u) = \text{essential}\sup_{0\leqslant u<2\pi} g(u) = \sup_{|z|<1} |F(z)|.$$

Solution by the proposer. Let C_r denote the circle with radius r and center 0. By Cauchy's integral formula, for n a positive integer and $|z| < r < 1$,

$$\begin{aligned} |F(z)|^n &= \left|(2\pi i)^{-1}\int_{C_r} F(v)^n (v-z)^{-1}\,dv\right| \\ &\leqslant (2\pi)^{-1}(r-|z|)^{-1}\int_0^{2\pi}|F(re^{iu})|^n\,du. \end{aligned}$$

So

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1}\liminf_{r\to 1-}\int_0^{2\pi}|F(re^{iu})|^n\,du.$$

Now $|F(re^{iu})|^n \leqslant h(u)^n$ for all $r<1$, and it follows from the Lebesgue dominated convergence theorem that

$$\liminf_{r\to 1-}\int_0^{2\pi}|F(re^{iu})|^n\,du \leqslant \int_0^{2\pi} g(u)^n\,du.$$

Thus

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1}\int_0^{2\pi} g(u)^n\,du \leqslant (1-|z|)^{-1}\left(\operatorname*{ess.\,sup}_{0\leqslant u<2\pi} g(u)\right)^n$$

and $|F(z)| \leqslant (1-|z|)^{-1/n}(\text{ess.}\sup_{0\leqslant u<2\pi} g(u))$. Let $n\to\infty$ to obtain $|F(z)| \leqslant \operatorname*{ess.\,sup}_{0\leqslant u<2\pi} g(u)$. The rest is clear.

Tangents to a Circle and a Parabola

6364 [1981, 711]. *Proposed by the late K. B. Leisenring, University of Michigan*

A circle with center at the vertex and radius equal to the latus rectum meets a parabola at P, Q. The circle and parabola have common tangents meeting the parabola at X, Y. Prove that XP, YQ are tangent to the circle.

Solution by Anders Bager, Akelejevej 5, DK-9800 Hjørring, Denmark. The parabola $y^2 = x$ and the circle $x^2+y^2=1$ meet at $P=(a,\sqrt{a})$ and $Q=(a,-\sqrt{a})$ where $a=(\sqrt{5}-1)/2$. The tangent at the point (t^2,t) on the parabola has equation $2ty = x+t^2$, and hence its distance from the origin 0 is $t^2(4t^2+1)^{-1/2}$. Setting this distance equal to 1 produces $X=(b,-\sqrt{b})$ and $Y=(b,\sqrt{b})$ where $b=2+\sqrt{5}$. Hence $\overrightarrow{OP}\cdot\overrightarrow{OX}=1$ and so

$$\overrightarrow{OP}\cdot\overrightarrow{PX} = \overrightarrow{OP}\cdot\left(\overrightarrow{OX}-\overrightarrow{OP}\right) = 1-1=0.$$

Thus PX is tangent to the circle and, similarly, so also is YQ.

Also solved by 21 other readers. The problem turned out to be rather elementary.

ANSWER TO PHOTO ON PAGE 330

It isn't often that four heads of the same department (at different times) can stand next to each other and smile. From left to right: Heini Halberstam, Paul Bateman, Stewart Cairns, and Mahlon Day. The department is at the University of Illinois; the order of seniority is Cairns, Day, Bateman, and Halberstam (the present head). The picture was taken by Robert J. McEliece.

$$\sup_{0 \leqslant u < 2\pi} g(u) = \underset{0 \leqslant u < 2\pi}{\text{essential} \sup} g(u) = \sup_{|z|<1} |F(z)|.$$

Solution by the proposer. Let C_r denote the circle with radius r and center 0. By Cauchy's integral formula, for n a positive integer and $|z| < r < 1$,

$$|F(z)|^n = \left|(2\pi i)^{-1} \int_{C_r} F(v)^n (v-z)^{-1} dv\right|$$
$$\leqslant (2\pi)^{-1}(r - |z|)^{-1} \int_0^{2\pi} |F(re^{iu})|^n du.$$

So

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1} \liminf_{r \to 1-} \int_0^{2\pi} |F(re^{iu})|^n du.$$

Now $|F(re^{iu})|^n \leqslant h(u)^n$ for all $r < 1$, and it follows from the Lebesgue dominated convergence theorem that

$$\liminf_{r \to 1-} \int_0^{2\pi} |F(re^{iu})|^n du \leqslant \int_0^{2\pi} g(u)^n du.$$

Thus

$$|F(z)|^n \leqslant (2\pi)^{-1}(1-|z|)^{-1} \int_0^{2\pi} g(u)^n du \leqslant (1-|z|)^{-1} \Big(\underset{0 \leqslant u < 2\pi}{\text{ess.}\sup} g(u)\Big)^n$$

and $|F(z)| \leqslant (1-|z|)^{-1/n}(\text{ess.}\sup_{0 \leqslant u < 2\pi} g(u))$. Let $n \to \infty$ to obtain $|F(z)| \leqslant \underset{0 \leqslant u < 2\pi}{\text{ess.}\sup} g(u)$. The rest is clear.

Tangents to a Circle and a Parabola

6364 [1981, 711]. *Proposed by the late K. B. Leisenring, University of Michigan*

A circle with center at the vertex and radius equal to the latus rectum meets a parabola at P, Q. The circle and parabola have common tangents meeting the parabola at X, Y. Prove that XP, YQ are tangent to the circle.

Solution by Anders Bager, Akelejevej 5, DK-9800 Hjørring, Denmark. The parabola $y^2 = x$ and the circle $x^2 + y^2 = 1$ meet at $P = (a, \sqrt{a})$ and $Q = (a, -\sqrt{a})$ where $a = (\sqrt{5} - 1)/2$. The tangent at the point (t^2, t) on the parabola has equation $2ty = x + t^2$, and hence its distance from the origin 0 is $t^2(4t^2 + 1)^{-1/2}$. Setting this distance equal to 1 produces $X = (b, -\sqrt{b})$ and $Y = (b, \sqrt{b})$ where $b = 2 + \sqrt{5}$. Hence $\overrightarrow{OP} \cdot \overrightarrow{OX} = 1$ and so

$$\overrightarrow{OP} \cdot \overrightarrow{PX} = \overrightarrow{OP} \cdot (\overrightarrow{OX} - \overrightarrow{OP}) = 1 - 1 = 0.$$

Thus PX is tangent to the circle and, similarly, so also is YQ.

Also solved by 21 other readers. The problem turned out to be rather elementary.

ANSWER TO PHOTO ON PAGE 330

It isn't often that four heads of the same department (at different times) can stand next to each other and smile. From left to right: Heini Halberstam, Paul Bateman, Stewart Cairns, and Mahlon Day. The department is at the University of Illinois; the order of seniority is Cairns, Day, Bateman, and Halberstam (the present head). The picture was taken by Robert J. McEliece.

Exponential Series

6365 [1981, 711]. *Proposed by Alan Wilde, University of Michigan.*

For integers $n > h \geqslant 0$, define $\exp_h(z) = \Sigma z^k/k!$, where the sum is over all nonnegative integers $k \equiv h(\bmod n)$.

(i) Assuming Fermat's Last Theorem, show that if $n > 2$, there is no complex number z such that $\exp_0(z)$ and $\exp_1(z)$ are nonzero rational numbers while $\exp_2(z) = \cdots = \exp_{n-1}(z) = 0$.

(ii)* Is Fermat's Last Theorem needed?

Solution by R. G. E. Pinch, Mathematical Institute, University of Oxford. The following two theorems appear in A. Baker, *Transcendental Number Theory*, Cambridge University Press, 1975.

THEOREM 12.1. *If both ξ_1, ξ_2 and η_1, η_2 are linearly independent over $\mathbb{Q}$, then two at least of $\xi_i, \eta_j, \exp(\xi_i\eta_j)$ $(i, j = 1,2)$ are transcendental.*

THEOREM 1.4. *If $\alpha_1,\ldots, \alpha_k$ are distinct algebraic numbers and $\beta_1,\ldots, \beta_k$ are algebraic numbers, not all zero, then*

$$\sum_{i=1}^{k} \beta_i \exp(\alpha_i) \neq 0.$$

The problem is now solved by the following theorem, the proof of which does not require Fermat's Last Theorem to be assumed.

THEOREM . *If $n > 2$ and $z \neq 0$, then $\exp_0(z),\ldots, \exp_{n-1}(z)$ cannot all be algebraic.*

Proof. Suppose they are. Then the numbers

$$\exp(w^i z) = \sum_{h=0}^{n-1} w^{hi}\exp_h(z) \quad (i = 0,1,2,\ldots),$$

where w is a primitive nth root of unity, are all algebraic as well.

Firstly, z cannot be a rational multiple of w, for if so, z and $\exp(z)$ are both algebraic, which contradicts Theorem 1.4.

But now, $\xi_1 = 1$, $\xi_2 = w$ and $\eta_1 = w$, $\eta_2 = z$ are linearly independent over $\mathbb{Q}$ and so by Theorem 12.1 at least two of $z, w, \exp(wz)$ are transcendental, a contradiction.

Also solved by Robert Breusch, E. G. Straus, William C. Waterhouse, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

13 Lectures on Fermat's Last Theorem. By Paulo Ribenboim. Springer-Verlag, New York, 1979. xvi + 302 pp.

DAVID R. HAYES
Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01002

While browsing in an antique store one afternoon, I came across a collection of old comic books. Among them I found one entitled "Donald Duck in Mathmagic Land." As I recall, the

Exponential Series

6365 [1981, 711]. *Proposed by Alan Wilde, University of Michigan.*

For integers $n > h \geqslant 0$, define $\exp_h(z) = \Sigma z^k/k!$, where the sum is over all nonnegative integers $k \equiv h (\mathrm{mod}\ n)$.

(i) Assuming Fermat's Last Theorem, show that if $n > 2$, there is no complex number z such that $\exp_0(z)$ and $\exp_1(z)$ are nonzero rational numbers while $\exp_2(z) = \cdots = \exp_{n-1}(z) = 0$.

(ii)* Is Fermat's Last Theorem needed?

Solution by R. G. E. Pinch, Mathematical Institute, University of Oxford. The following two theorems appear in A. Baker, *Transcendental Number Theory*, Cambridge University Press, 1975.

THEOREM 12.1. *If both ξ_1, ξ_2 and η_1, η_2 are linearly independent over $\mathbb{Q}$, then two at least of $\xi_i, \eta_j, \exp(\xi_i\eta_j)$ $(i, j = 1,2)$ are transcendental.*

THEOREM 1.4. *If $\alpha_1, \ldots, \alpha_k$ are distinct algebraic numbers and $\beta_1, \ldots, \beta_k$ are algebraic numbers, not all zero, then*

$$\sum_{i=1}^{k} \beta_i \exp(\alpha_i) \neq 0.$$

The problem is now solved by the following theorem, the proof of which does not require Fermat's Last Theorem to be assumed.

THEOREM . *If $n > 2$ and $z \neq 0$, then $\exp_0(z), \ldots, \exp_{n-1}(z)$ cannot all be algebraic.*

Proof. Suppose they are. Then the numbers

$$\exp(w^i z) = \sum_{h=0}^{n-1} w^{hi} \exp_h(z) \quad (i = 0,1,2,\ldots),$$

where w is a primitive nth root of unity, are all algebraic as well.

Firstly, z cannot be a rational multiple of w, for if so, z and $\exp(z)$ are both algebraic, which contradicts Theorem 1.4.

But now, $\xi_1 = 1$, $\xi_2 = w$ and $\eta_1 = w$, $\eta_2 = z$ are linearly independent over $\mathbb{Q}$ and so by Theorem 12.1 at least two of $z, w, \exp(wz)$ are transcendental, a contradiction.

Also solved by Robert Breusch, E. G. Straus, William C. Waterhouse, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

13 Lectures on Fermat's Last Theorem. By Paulo Ribenboim. Springer-Verlag, New York, 1979. xvi + 302 pp.

DAVID R. HAYES
Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01002

While browsing in an antique store one afternoon, I came across a collection of old comic books. Among them I found one entitled "Donald Duck in Mathmagic Land." As I recall, the

story began with Donald in a sweat over the disarray of his personal finances. Most troublesome was a large loan from his Uncle Scrooge which must be repaid soon. Suddenly a caped figure appeared, the *Spirit of Mathematics*, and whisked Donald away to a magic land where the flora and fauna were twisted into geometric and numerical shapes. Although I don't now remember seeing it, somewhere in that surrealistic landscape there must have been an irregular prime chirping from its nest in a Bernoulli number.

You can discover your own wide path to Mathmagic Land in these lectures of Ribenboim. From the moment he states Fermat's Last "Theorem,"

> *The equation $x^n + y^n = z^n$, where n is a natural number greater than* 2, *has no solution in integers all different from* 0,

in the first lecture, Ribenboim provides the reader with a varied, amusing and enlightening flow of information. Did you know that Fermat's copy of the complete works of Diophantus with the most famous margin in scientific history is lost and that we must depend upon a secondary source for the quote, "... I have discovered a truly remarkable proof which this margin is too small to contain?" Did you know that if a counterexample (x, y, z) to FLT exists, then x must be magnitudes larger than the number of atomic nuclei which could be packed into the Universe? These two bits of information are sampled from the historical survey in Lecture I and the statement of the latest results on FLT in Lecture II. Both these initial lectures are accessible to anyone who remembers his high school mathematics.

In Lecture III, entitled *B.K. = Before Kummer*, Ribenboim reproduces the proof given by Fermat himself for the case $n = 4$. After $n = 4$, one need only consider $n = p$, where p is an odd prime, and Ribenboim more or less restricts himself to that case in the subsequent lectures. He concludes Lecture III by settling FLT for the exponents $n = p = 3$, 5 and 7.

The next few lectures form the heart of the book. Beginning with *The Naive Approach* (Lecture IV), Ribenboim proceeds to survey the fruits of three centuries of intense effort. Here it all is: The astounding array of necessary conditions which must be satisfied if there is a solution in which p does not divide xyz, the so-called "first case" of FLT; the surprising connections with Bernoulli numbers which were spaded out of the problem by the Herculean labors of Kummer; and the extensive numerical calculations carried out first by hand and then, in modern times, electronically, culminating in the recent results of Wagstaff which verify the truth of FLT for all odd prime exponents less than 125,000. The Fermat Problem has clearly absorbed an enormous amount of intellectual capital!

A naive reader, floating along on this flood of results, may well begin to feel lost in a mathmagical fog. Ribenboim does not directly address questions like, "Why is FLT important?" or, "Where is all this leading?" Indeed, somewhere in the middle of Lecture V, the naive reader may conclude that people are interested in FLT mainly as a sport. Ribenboim's style tends to reinforce this idea. For example, in Lecture VIII, as though describing the appearance of a new slugger in a history of baseball, Ribenboim writes: "Then came Wieferich ... Few people were able to understand how Wieferich succeeded, like a magician, in unravelling from very complicated formulas, so simple and beautiful a criterion as:

> *If the first case fails for the exponent p, then p must satisfy the congruence*
>
> $$2^{p-1} \equiv 1 \pmod{p^2}.\text{"}$$

By the way, there are many other names in the FLT Hall of Fame, all conveniently listed with their papers in extensive bibliographies at the end of *each* lecture.

In his *A Mathematician's Apology*, G. H. Hardy attempts to define a deep sense in which a theorem or problem is "serious." In Lectures V–VII, Ribenboim describes *Kummer's Monument*, the work which in the middle 1800's penetrated deeply into the problem and in the process founded the theory of algebraic numbers. The committed reader, naive or not, cannot complete

these three lectures without feeling that he has passed through some profound depths. Like a maturing love affair, the Fermat Problem suddenly becomes "serious." The level of exposition in these lectures is very high. The author skillfully sketches the necessary background in the theory of cyclotomic fields, providing an understanding of Kummer's ideas and techniques which is accessible to a wide audience. Ribenboim does not give complete proofs of the major theorems, but he does give a detailed outline of how the proofs go and of the difficulties which Kummer faced and overcame.

In Lecture IX, Ribenboim indicates how class field theory can be used to improve both the results he has presented in previous lectures and one's understanding of why they are true. Even the reader without background in class field theory can gain insight from this lecture. Ribenboim sketches modern developments related to FLT in Lecture X. In particular, he mentions the Iwasawa theory and the connections to points of finite order on elliptic curves. In Lecture XII and the amusing Lecture XIII, *Variations and Fugue on a Theme*, Ribenboim reports on what happens when mathematicians invoke their license to modify a problem when they cannot solve it. For example, "Variation I (In the Tone of Polynomial Functions)" considers solutions to the Fermat equation in the ring of polynomials over a base field K. In work too recent to be included in these lectures, David Goss has shown that when K is finite, there is a much more profound analogue of the Fermat Problem than the obvious one. Goss' analogue has most of the features we have come to associate with FLT itself.

On his return from Mathmagic Land, Donald Duck offered his Uncle Scrooge a deal. He would work for Scrooge for thirty days. The first day, his wage would be one cent, and every day thereafter his wage would be double that of the previous day. Uncle Scrooge bought it. It's clear that Donald profited from his trip to Mathmagic Land! You also will profit from reading this book of Ribenboim, but the coinage you receive will not be stamped out by the U.S. Mint.

The Tragicomical History of Thermodynamics 1822-1854. By C. Truesdell. Springer-Verlag, New York, 1980. xii + 372 pp.

Stuart S. Antman
Department of Mathematics, University of Maryland, College Park, MD 20742

As an undergraduate I had the misfortune to study thermodynamics after studying advanced calculus. My course in thermodynamics treated the state of a given body of gas as described by its pressure p, temperature θ, and density ρ when these variables do not vary with position and when they change so slowly ("quasistatically") with time that their time derivatives are negligible. The differential as an infinitesimal, justly banished from analysis, was the main tool used for describing such processes: The differential was the time derivative of a function independent of time. (There were in fact several species of such differentials, one denoted by a "d" bearing a cross.) The fundamental laws of thermodynamics told how the gas would behave in a heat engine having the grace to preserve the spatial uniformity of the gas while coaxing it along quasistatically. The existence of such an ideal engine was never questioned, although a version of the Second Law of Thermodynamics did prohibit another such ideal engine, the perpetual motion machine. The quasistatic evolution of the gas was expressed in terms of differentials of p, θ, ρ, differentials of other quantities such as internal energy, entropy, enthalpy, and free energy (which were functionally related to each other and to p, θ, ρ), and differentials of quantities such as the heat, which doesn't have a differential. Each function relating one set of thermodynamical variables to another having the same cardinality was assumed smooth and invertible (according to what Truesdell calls the "Principle of Thoughtless Invertibility" (p. 283)). We students were required to compute all

these three lectures without feeling that he has passed through some profound depths. Like a maturing love affair, the Fermat Problem suddenly becomes "serious." The level of exposition in these lectures is very high. The author skillfully sketches the necessary background in the theory of cyclotomic fields, providing an understanding of Kummer's ideas and techniques which is accessible to a wide audience. Ribenboim does not give complete proofs of the major theorems, but he does give a detailed outline of how the proofs go and of the difficulties which Kummer faced and overcame.

In Lecture IX, Ribenboim indicates how class field theory can be used to improve both the results he has presented in previous lectures and one's understanding of why they are true. Even the reader without background in class field theory can gain insight from this lecture. Ribenboim sketches modern developments related to FLT in Lecture X. In particular, he mentions the Iwasawa theory and the connections to points of finite order on elliptic curves. In Lecture XII and the amusing Lecture XIII, *Variations and Fugue on a Theme*, Ribenboim reports on what happens when mathematicians invoke their license to modify a problem when they cannot solve it. For example, "Variation I (In the Tone of Polynomial Functions)" considers solutions to the Fermat equation in the ring of polynomials over a base field K. In work too recent to be included in these lectures, David Goss has shown that when K is finite, there is a much more profound analogue of the Fermat Problem than the obvious one. Goss' analogue has most of the features we have come to associate with FLT itself.

On his return from Mathmagic Land, Donald Duck offered his Uncle Scrooge a deal. He would work for Scrooge for thirty days. The first day, his wage would be one cent, and every day thereafter his wage would be double that of the previous day. Uncle Scrooge bought it. It's clear that Donald profited from his trip to Mathmagic Land! You also will profit from reading this book of Ribenboim, but the coinage you receive will not be stamped out by the U.S. Mint.

The Tragicomical History of Thermodynamics 1822-1854. By C. Truesdell. Springer-Verlag, New York, 1980. xii + 372 pp.

STUART S. ANTMAN
Department of Mathematics, University of Maryland, College Park, MD 20742

As an undergraduate I had the misfortune to study thermodynamics after studying advanced calculus. My course in thermodynamics treated the state of a given body of gas as described by its pressure p, temperature θ, and density ρ when these variables do not vary with position and when they change so slowly ("quasistatically") with time that their time derivatives are negligible. The differential as an infinitesimal, justly banished from analysis, was the main tool used for describing such processes: The differential was the time derivative of a function independent of time. (There were in fact several species of such differentials, one denoted by a "d" bearing a cross.) The fundamental laws of thermodynamics told how the gas would behave in a heat engine having the grace to preserve the spatial uniformity of the gas while coaxing it along quasistatically. The existence of such an ideal engine was never questioned, although a version of the Second Law of Thermodynamics did prohibit another such ideal engine, the perpetual motion machine. The quasistatic evolution of the gas was expressed in terms of differentials of p, θ, ρ, differentials of other quantities such as internal energy, entropy, enthalpy, and free energy (which were functionally related to each other and to p, θ, ρ), and differentials of quantities such as the heat, which doesn't have a differential. Each function relating one set of thermodynamical variables to another having the same cardinality was assumed smooth and invertible (according to what Truesdell calls the "Principle of Thoughtless Invertibility" (p. 283)). We students were required to compute all

possible partial derivatives of these functions, which was easy because we could invoke the related principle that "No Jacobian ever vanishes," apparently the true First Law of Thermodynamics. But the pleasure of studying a science exempt from the rigors of analysis was scant compensation for the absence of principles as clean as "$\mathbf{f} = m\mathbf{a}$" and the absence of problems as substantial as those of classical mechanics.

It was consequently with relief and satisfaction that I read Truesdell's telling critiques [**1**], [**2**] of the incomprehensibility of expositions of thermodynamics. His prescription for the surfeit of differentials was simple: Variables such as θ are to be considered as functions of position and time. Then expressions containing the differential $d\theta$ can be replaced by precise and illuminating expressions containing the partial derivatives of θ. For example, a typical contemporary text might state that the heat δQ added to a spatially uniform gas in a quasistatic process taking (θ, ρ) to $(\theta + d\theta, \rho + d\rho)$ is

$$\delta Q = A(\theta, \rho)\,d\theta + B(\theta, \rho)\,d\rho \tag{1}$$

where the functions A and B characterize the thermomechanical properties of the gas. (The δ before Q in (1) could be replaced with a cross-bearing d.) The vague expression (1) can be replaced by a statement that the local rate per unit volume at which heat is added to the material at $\mathbf{x}$ at time t when θ and ρ are specified functions of $\mathbf{x}$ and t is simply

$$a(\theta(\mathbf{x}, t),\quad \rho(\mathbf{x}, t))\theta_t(\mathbf{x}, t) + b(\theta(\mathbf{x}, t),\quad \rho(\mathbf{x}, t))\rho_t(\mathbf{x}, t) \tag{2}$$

where a and b are proportional to A and B and where the subscripts denote partial derivatives. The total heat added to a body of gas in a process lasting from t_1 to t_2 is just the integral of (2) over (t_1, t_2) and over the set of $\mathbf{x}$'s forming the material points of the gas. The use of (2) not only conforms exactly to the notion of rate, but also makes it clear that thermodynamics need not be chained to quasistatic processes for bodies described by spatially constant fields. In general, (2) is not the partial time derivative of any function with values $q(\theta(\mathbf{x}, t), \rho(\mathbf{x}, t))$, i.e., the differential 1-form on the right side of (1) fails to be exact. Consequently the appearance of Q on the left side of (1) is misleading. (The use of such notations is a source of the difficulties many students of science experience with the concept of virtual work. The notion of exactness of forms plays a central role in the traditional theory of entropy.)

Truesdell's prescription for the murkiness of the statements of the fundamental laws was likewise straightforward: The laws should be expressed as mathematical relations among functions describing physical entities and not as rules for the operation of hypothetical steam engines. In this way their validity could be determined by scientific test rather than by speculation almost theological in nature. In these writings Truesdell promoted the adoption of a language for civilized discourse in a thermodynamics rich enough to encompass fields varying in space and time. This language is just that of clean mathematics. (The determination of the correct versions of well formulated laws of continuum thermomechanics and of their consequences is an open problem which is being actively pursued.)

How did it happen that thermodynamics seems to be so conceptually inferior to continuum mechanics and electromagnetic theory, the two other classical field theories of physics? The book under review, its title suggesting both entertainment and narrowly focused historical analysis, offers convincing answers to this question.

The subject of Truesdell's history is the conceptual founding of thermodynamics primarily by Carnot, Clausius, and Reech, and to a lesser extent by Rankine, Kelvin, and others. (Fourier's work, duly analyzed, contributed little to this process: His *Analytic Theory of Heat* did not treat heat.) Truesdell blames the scientific inferiority of early thermodynamics to early mechanics on the mathematical inferiority of these founders of thermodynamics (except Kelvin) relative to the founders of mechanics, namely Newton, the Bernoullis, Euler, and Cauchy: "The vagueness and vacillation of Carnot's concepts and assumptions... are typical of a theory not subjected to the discipline of mathematical statement, and perhaps unavoidable in such a theory" (p. 101).

"Among physicists of the first rank, Carnot is the first who was not in at least equal measure a mathematician. Thermodynamics is the first mathematical science to have been invented without the control afforded by patient, merciless, mathematical criticism. It has suffered from this congenital defect from 1824 until now" (p. 136). "Obscure logic, the painfully awkward calculus that dogs standard thermodynamics to this day, and vague expression have joined Clausius' splendid achievements to form his legacy" (p. 207). Truesdell attributes the continued inferiority of thermodynamics to mechanics to the slavish adherence of the successors of Carnot and Clausius to the unfortunate expository precedent they established. Even Kelvin and Poincaré could not shake the dead hand of thermodynamics' sorry mathematical tradition (cf. pp. 143, 168).

Truesdell's book contains three kinds of exposition: (i) straightforward accounts of the primary literature in a unified notation, (ii) a lively set of strongly opinionated (and clearly identified) critiques, and (iii) retrospective sketches of what could have been achieved in the early history of thermodynamics with the tools then available. The "ahistorical moralizing" (p. 5) of the critiques make the book consistently entertaining: "... Carnot does not follow the tradition of eighteenth-century rational mechanics he has just praised for its generality and extent. Instead, the sardonic muse directs him to write in a medium that anybody can understand. An obvious necessary condition is that no mathematics be used in the main text. This condition did not turn out to be sufficient.... He is reported to have insisted that his brother, untrained in the subject, read and criticize the work; according to the legend, his brother understood it perfectly. Later students, unable to seek help from that brother, have puzzled, are puzzling, and forever will puzzle over it" (p. 80). The same pungency can be found in the amusingly apt quotations from the *Divine Comedy* that introduce each chapter; these are worth translating. Who could resist a book with a section entitled: "The Disastrous Effects of Experiment on the Development of Thermodynamics, 1812-1853"?

One can argue that it is not fair to compare nineteenth-century thermodynamics invidiously with eighteenth-century mechanics: The pioneers of thermodynamics should rather be judged against the pioneers of mechanics who preceded Newton. Such a view ignores the rich arsenal of mathematical technique available to the early thermodynamicists but not to Newton and his predecessors. Moreover, the pioneers of thermodynamics had before them the exemplar of a mature and coherent mathematical science, mechanics, whereas the pioneers of mechanics had to overcome a two thousand year old tradition of disputatious Aristotelianism. But the telling objection to this claim of unfairness is that the founders of electromagnetic theory faced the same obstacles as the founders of thermodynamics at the same time, yet succeeded in creating a coherent and beautiful mathematical science.

One need only examine the progression from Newtonian to Lagrangean to Hamiltonian mechanics as presented in most contemporary texts on discrete mechanics to realize that the logical standards of traditional expositions of other fields of classical physics cannot withstand careful scrutiny. But the most general and most accessible formulation of classical mechanics, the Newtonian, admits a perfectly clear presentation, while the logical deficiencies in the introduction of the other formulations can be removed by the same methods that remove them in thermodynamics. Newtonian mechanics also encompasses innumerable specific, tractable, and illuminating problems. Nineteenth-century thermodynamics had nothing comparable to it. (I surmise that thermodynamics got off to such a bad start because Carnot, influenced by the Lagrangean formalism, modeled thermodynamics on the statics of discrete systems described by the Principle of Virtual Work expressed in terms of the ambiguous differentials, rather than on Euler's more appropriate field theory of perfect fluids.)

Can a reader untrained in classical thermodynamics follow the scientific development of the book without serious difficulty? I think so. By keeping one eye on the critiques following the straight historical analysis, such a reader can avoid being upset by failing to follow an argument later criticized as faulty. Alternatively, this history could serve as a vademecum to anyone wishing to assault a traditional exposition.

Of what value is this specialized historical tract to a mathematician, one, say, who is not consumed with an overwhelming desire to follow the involutions of the genesis of a physical theory that has had little impact on mathematics? There are several answers: The book gives a vivid impression of the struggles necessary to establish a mathematical science describing physical reality. Thereby it indirectly gives an appreciation of the genius of such as Euler, Cauchy, and Maxwell, who triumphed in their struggles to establish continuum mechanics and electromagnetism. It contains the evaluations by one who is not only a historian of science, but is also an active researcher who has made significant contributions to thermodynamics. It makes plain that the problems of cleaning up the exposition of thermodynamics and distinguishing principle from approximation require not the full array of modern mathematical technique, but only precisely stated hypotheses and the careful language of advanced calculus. To mathematicians as purveyors of the language of science it offers the disturbing suggestion that there is some law of inertia by which authors of contemporary books in the mathematical sciences adhere to clumsy mathematical formulations because their predecessors, from whom they uncritically copied, used those very formulations.

Why don't scientists approach their disciplines with a critical mind? A few do. (Indeed, Truesdell protects his flank from attack by physicists with quotes by the Nobel laureate P. W. Bridgman on the deficiencies of thermodynamics.) But most of the scientists and engineers we educate in this country, who typically take little more than two years of college mathematics and who, never having seen a logically coherent presentation of a mathematical science, regard science as a bunch of memorized procedures, are simply unequipped to detect mathematical or logical weaknesses. Most mathematicians attribute their failure to comprehend a typical presentation of a physical subject to their lack of physical intuition rather than to the inadequacy of the presentation. What we need is an educational system creating scientists so secure in their knowledge that they can (justly) say, "If I can't understand it, then it's incomprehensible."

References

1. C. Truesdell and R. A. Toupin, The Classical Field Theories, Chapter E, Handbuch der Physik III/1, edited by S. Flügge, Springer-Verlag, Berlin, Göttingen, Heidelberg, 1960.

2. C. Truesdell, Rational Thermodynamics, A Course of Lectures on Selected Topics, McGraw-Hill, New York, 1969.

MISCELLANEA

104. Sir James Jeans went so far as to assert that the Almighty must be a mathematician Himself, or otherwise natural phenomena could not be described in such beautiful mathematical terms. This argument takes us rather out of our depth, but we hope we shall not be accused of impiety if we remark that it is clear that no such feeling for beauty inspires the Board of Inland Revenue (in the U.S.A., the Internal Revenue Service—Editor).

—B. V. BOWDEN, Faster Than Thought, London, 1953, p. 259

105. The mathematicians are well acquainted with the difference between pure science, which has to do only with ideas, and the applications of its laws to the use of life, in which they are constrained to submit to the imperfection of matter and the influence of accidents.

—SAMUEL JOHNSON, *The Rambler*, no. 14, May 5, 1750

Of what value is this specialized historical tract to a mathematician, one, say, who is not consumed with an overwhelming desire to follow the involutions of the genesis of a physical theory that has had little impact on mathematics? There are several answers: The book gives a vivid impression of the struggles necessary to establish a mathematical science describing physical reality. Thereby it indirectly gives an appreciation of the genius of such as Euler, Cauchy, and Maxwell, who triumphed in their struggles to establish continuum mechanics and electromagnetism. It contains the evaluations by one who is not only a historian of science, but is also an active researcher who has made significant contributions to thermodynamics. It makes plain that the problems of cleaning up the exposition of thermodynamics and distinguishing principle from approximation require not the full array of modern mathematical technique, but only precisely stated hypotheses and the careful language of advanced calculus. To mathematicians as purveyors of the language of science it offers the disturbing suggestion that there is some law of inertia by which authors of contemporary books in the mathematical sciences adhere to clumsy mathematical formulations because their predecessors, from whom they uncritically copied, used those very formulations.

Why don't scientists approach their disciplines with a critical mind? A few do. (Indeed, Truesdell protects his flank from attack by physicists with quotes by the Nobel laureate P. W. Bridgman on the deficiencies of thermodynamics.) But most of the scientists and engineers we educate in this country, who typically take little more than two years of college mathematics and who, never having seen a logically coherent presentation of a mathematical science, regard science as a bunch of memorized procedures, are simply unequipped to detect mathematical or logical weaknesses. Most mathematicians attribute their failure to comprehend a typical presentation of a physical subject to their lack of physical intuition rather than to the inadequacy of the presentation. What we need is an educational system creating scientists so secure in their knowledge that they can (justly) say, "If I can't understand it, then it's incomprehensible."

References

1. C. Truesdell and R. A. Toupin, The Classical Field Theories, Chapter E, Handbuch der Physik III/1, edited by S. Flügge, Springer-Verlag, Berlin, Göttingen, Heidelberg, 1960.

2. C. Truesdell, Rational Thermodynamics, A Course of Lectures on Selected Topics, McGraw-Hill, New York, 1969.

MISCELLANEA

104. Sir James Jeans went so far as to assert that the Almighty must be a mathematician Himself, or otherwise natural phenomena could not be described in such beautiful mathematical terms. This argument takes us rather out of our depth, but we hope we shall not be accused of impiety if we remark that it is clear that no such feeling for beauty inspires the Board of Inland Revenue (in the U.S.A., the Internal Revenue Service—Editor).

—B. V. Bowden, Faster Than Thought, London, 1953, p. 259

105. The mathematicians are well acquainted with the difference between pure science, which has to do only with ideas, and the applications of its laws to the use of life, in which they are constrained to submit to the imperfection of matter and the influence of accidents.

—Samuel Johnson, *The Rambler*, no. 14, May 5, 1750

TWO-YEAR COLLEGE MATHEMATICS READINGS

Edited by Warren Page

viii + 304 pages. Paperbound
List $19.50 MAA Member: $14.50

This work surveys the two-year college mathematics landscape. Stops are made at points of interest that cover areas from algebra to recreational mathematics.

There are many things of interest for everyone in this book. Some are expository articles that introduce the reader to new advances in mathematics. Other articles can be used to enrich course work and to stimulate both the student's and instructor's mathematical intellect.

This book will serve as an excellent source of enrichment for the talented student. Teachers of undergraduates will find it is an excellent reference source that they will want to consult when preparing lectures.

Topics covered are:

- **Algebra**
- **Geometry**
- **Number Theory**
- **Calculus**
- **Probability and Statistics**
- **Calculators and Computers**
- **Mathematics Education**
- **Recreational Mathematics**

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 6 June–July 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

ON THE GEOMETRY OF THE KEPLER PROBLEM

JOHN MILNOR
Institute for Advanced Study, Princeton NJ 08540

Abstract. It will be convenient to use the term *Kepler orbit* for any curve $\mathbf{x} = \mathbf{x}(t)$ in 3-space which arises as a solution to the Newtonian two body problem. Hamilton showed that the velocity vector $\mathbf{v} = d\mathbf{x}/dt$, associated with any nondegenerate Kepler orbit, moves along a circle. Following Györgyi, Moser, Osipov and Belbruno, this *velocity circle* can be interpreted as follows. If we fix the total energy E, then the manifold M_E consisting of all vectors $\mathbf{v}$ with $\mathbf{v} \cdot \mathbf{v} > 2E$ possesses a natural Riemannian metric of constant curvature $-2E$, whose geodesics are precisely the circles associated in this way with Kepler orbits. In other words, M_E can be identified with one of the three classical geometries, that is with spherical, Euclidean or Lobachevsky space, so that each "straight line" in this geometry corresponds to a unique Kepler orbit.

1. The Velocity Circle. In Kepler's first attempts to understand the orbits of the planets, he tried the hypothesis that each orbit is a circle which lies in some plane containing the sun, but is not centered at the sun. This is of course wrong. Yet is does describe the correct answer to a slightly transformed problem.

Consider a particle which is attracted to the origin by a force inversely proportional to the square of the distance, so that its position vector $\mathbf{x} = \mathbf{x}(t)$ satisfies the *Newton differential equation*

$$d^2\mathbf{x}/dt^2 = -k\mathbf{x}/|\mathbf{x}|^3. \tag{1}$$

Here k is some fixed positive constant, and $|\mathbf{x}|$ denotes the Euclidean length $\sqrt{\mathbf{x} \cdot \mathbf{x}}$ of the vector $\mathbf{x}$. (This equation is related to the Newtonian two body problem as follows. If two spherical objects with position coordinates $\mathbf{x}_1$ and $\mathbf{x}_2$ move about their common center of gravity in accordance with Newton's laws, then the difference vector $\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$ will satisfy (1).) Solutions of this equation, whether elliptical or otherwise, will be called *Kepler orbits.*

Throughout §1, *we will assume that the position* $\mathbf{x}$ *and velocity* $\mathbf{v} = d\mathbf{x}/dt$ *are linearly independent vectors at time* $t = t_0$.

THEOREM 1 (Hamilton, 1846). *As t varies, the velocity vector $\mathbf{v} = d\mathbf{x}/dt$ moves along a circle C, which lies in some plane P containing the origin, but is not in general centered at the origin. Any such circle can occur, and this "velocity circle" C, together with its orientation, determines the orbit* $\mathbf{x} = \mathbf{x}(t)$ *uniquely.*

The proof will show that the corresponding orbit is either an ellipse, hyperbola, or parabola according as the origin lies inside, outside, or exactly on the velocity circle C. In the elliptic case, the velocity vector moves around the entire circle, but in the other two cases only that portion of the circle which is convex towards the origin is actually traversed. (Compare Fig. 1.)

To begin the proof, recall that the cross product vector $\mathbf{h} = \mathbf{x} \times \mathbf{v}$ is an invariant of the motion. That is the derivative $d\mathbf{h}/dt$ is identically zero; as one verifies by an easy calculation. Thinking of $\mathbf{x}(t)$ as the orbit of a particle of unit mass, we will call the length $h = |\mathbf{h}|$ the *angular momentum* of this orbit. Note that $h > 0$ by our linear independence hypothesis.

It will be convenient to introduce a new system of cartesian coordinates $\mathbf{x} = (x, y, z)$ *so that the vector* $\mathbf{h}$ *points along the positive z-axis.* The equation $\mathbf{x}(t) \times \mathbf{v}(t) = \mathbf{h} = (0, 0, h)$ then implies that the vector $\mathbf{x}(t)$ always lies in the (x, y)-plane. Setting $\mathbf{x}(t) = (r\cos\theta, r\sin\theta, 0)$, a straightforward

John Milnor received his Ph.D. at Princeton University in 1954, under the direction of Ralph Fox. As a student, he was also strongly influenced by Norman Steenrod, and by a year in Zürich under Heinz Hopf. After teaching for some years at Princeton University, interrupted by stays in Oxford, Berkeley, U.C.L.A.,and M.I.T., he moved across town to the Institute for Advanced Study. His mathematical work was centered around the topology of manifolds and related areas of algebra.

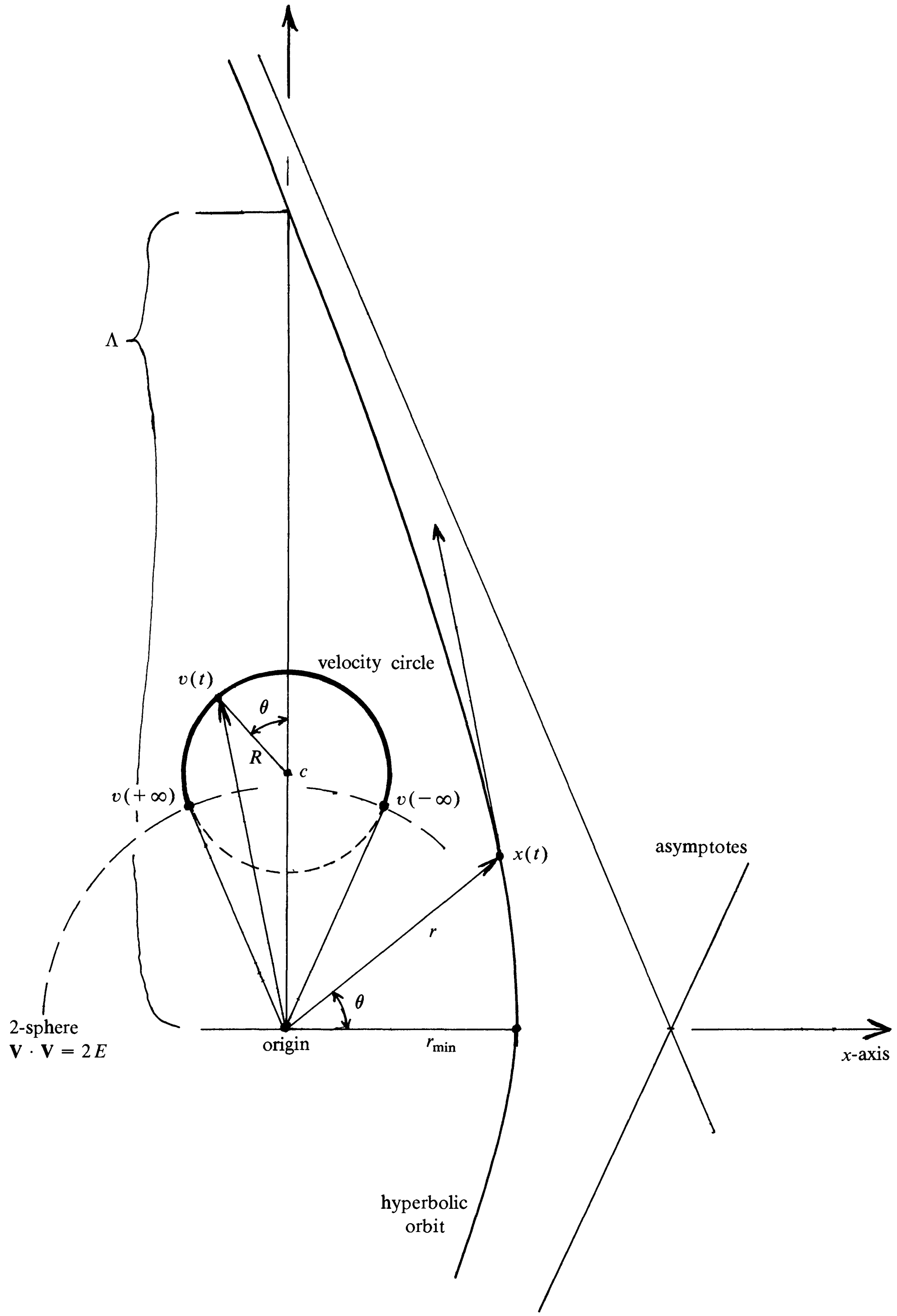

Λ
velocity circle
v(t)
θ
R
c
v(+∞)
v(−∞)
x(t)
asymptotes
r
θ
2-sphere
V · V = 2E
origin
r_min
x-axis
hyperbolic
orbit

FIG. 1

computation shows that

$$h = r^2\, d\theta/dt, \tag{2}$$

The constancy of this expression $r^2\, d\theta/dt$ is just *Kepler's Second Law*, which asserts that a line segment from the origin to $\mathbf{x}(t)$ traverses equal areas in equal times.

[It is interesting to note that Kepler's interpretation of this law was very different from our modern interpretation, since he had no concept of inertia. He believed that the sun exerts a sideways force, inversely proportional to distance, which pushes the planets around their orbits.]

Let us write the Newton equation as $d\mathbf{v}/dt = -k(\cos\theta, \sin\theta, 0)/r^2$. Dividing by $d\theta/dt = h/r^2$, and setting $R = k/h$, we obtain

$$d\mathbf{v}/d\theta = -R(\cos\theta, \sin\theta, 0),$$

Integrating, we obtain

$$\mathbf{v} = -R(\cos\theta, \sin\theta, 0)\, d\theta = R(-\sin\theta, \cos\theta, 0) + \mathbf{c},$$

where $\mathbf{c} = (c_1, c_2, 0)$ is a constant of integration. *This proves that the velocity vector* $\mathbf{v}$ *moves along a circle centered at* $\mathbf{c}$, *with radius* $R = k/h$ *inversely proportional to angular momentum, and lying in the plane through the origin which is orthogonal to the vector* $\mathbf{h}$.

It is interesting to note that the difference vector $\mathbf{v} - \mathbf{c} = R(-\sin\theta, \cos\theta, 0)$ is always orthogonal to the position vector $\mathbf{x} = r(\cos\theta, \sin\theta, 0)$. *Hence the angle between two velocity vectors, as measured around this circle C, is equal to the angle between corresponding position vectors as seen from the origin.*

The ratio $\varepsilon = |\mathbf{c}|/R$, that is the distance of the center from the origin divided by the radius, is called the *eccentricity* of the circle C with respect to the origin. *It will be convenient to choose our coordinates x, y for the plane so that the center* $\mathbf{c}$ *of the circle C lies on the positive y-axis.* With this convention, we can write

$$\mathbf{v} = R(-\sin\theta, \quad \varepsilon + \cos\theta, 0).$$

Substituting this expression in the equation $\mathbf{h} = \mathbf{x} \times \mathbf{v}$ we obtain the formula $h = rR(1 + \varepsilon\cos\theta)$ for angular momentum; and solving for r we obtain

$$r = \Lambda/(1 + \varepsilon\cos\theta), \tag{3}$$

where $\Lambda = h^2/k$. *This is precisely the equation of a conic section of eccentricity ε, with focus at the origin, in polar coordinates.* (See Appendix 1.) This conic section is either a circle, ellipse, parabola, or hyperbola according as $\varepsilon = 0$, $0 < \varepsilon < 1$, $\varepsilon = 1$, or $\varepsilon > 1$. In the latter two cases, note that the possible values of the angular coordinate θ are constrained by the inequality $1 + \varepsilon\cos\theta > 0$.

Our coordinate system has been chosen so that the point of closest approach, where $r = r_{\min}$, lies on the positive x-axis, with $\theta = 0$. This corresponds to the classical convention that the angular coordinate θ, known as the *anomaly*, should be measured from this point of closest approach. The geometrical constant Λ, proportional to the square of angular momentum, is known classically as the *semi-latus-rectum*. It can be described as the distance from origin to orbit in a direction at right angles to the direction of closest approach.

To complete the proof of Theorem 1, we must show that every such conic section (3) really yields a solution to the Newton equation. But Equation (3) shows that r can be expressed as a function of θ, and Equation (2) implies that $t = \int r(\theta)^2\, d\theta/h$ can also be expressed as a function of θ. It follows from the inverse function theorem that θ and hence $\mathbf{x}$ can be expressed as functions of t; and it is not difficult to check that the functions constructed in this way do indeed satisfy Equation (1). Details will be left to the reader. ∎

Another classical invariant associated with a solution $\mathbf{x} = \mathbf{x}(t)$ of Equation (1) is the *energy* $E = \mathbf{v}\cdot\mathbf{v}/2 - k/r$. A straightforward calculation shows that the derivative dE/dt is zero, so that E is indeed constant along any orbit. Note that the inequality $\mathbf{v}\cdot\mathbf{v} > 2E$ must always be satisfied.

In terms of the velocity circle, with radius R and center $\mathbf{c}$, energy is given by

$$2E = \mathbf{c}\cdot\mathbf{c} - R^2 = -(1-\varepsilon^2)k^2/h^2, \tag{4}$$

as one can check by evaluating E at any point of the orbit. This computation can be expressed more geometrically by the following statement, which is essentially due to Euclid. (Compare Coxeter [4] pp. 8, 81.)

LEMMA 1. *Let* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *be two points of the circle* C *which lie on a common line through the origin. Then* $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{c}\cdot\mathbf{c} - R^2$; *hence* $\mathbf{v}_1\cdot\mathbf{v}_2$ *is equal to* $2E$.

Proof. Let $\mathbf{w}$ be the point of the circle which is diametrically opposite to $\mathbf{v}_1$. Setting $\mathbf{v}_1 = \mathbf{c}+\mathbf{e}$ and $\mathbf{w} = \mathbf{c}-\mathbf{e}$, we see that

$$\mathbf{v}_1\cdot\mathbf{w} = \mathbf{c}\cdot\mathbf{c} - \mathbf{e}\cdot\mathbf{e} = \mathbf{c}\cdot\mathbf{c} - R^2 = 2E.$$

Since $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w}$ is a right triangle, it follows that $\mathbf{v}_1$ is orthogonal to the difference vector $\mathbf{v}_2 - \mathbf{w}$. Hence $\mathbf{v}_1\cdot\mathbf{v}_2$ is equal to $\mathbf{v}_1\cdot\mathbf{w} = 2E$, as required. ∎

In the case of an elliptical orbit, the precise shape and period of the orbit are related to the energy E as follows. First note that the *major axis* $2a$ of the ellipse is given by the computation

$$2a = r_{\min} + r_{\max} = \Lambda/(1+\varepsilon) + \Lambda/(1-\varepsilon) = 2\Lambda/(1-\varepsilon^2).$$

Since $\Lambda = h^2/k$, it follows from (4) that

$$2E = -k/a. \tag{5}$$

Thus the energy E *of an elliptic orbit is negative, and* $|E|$ *is inversely proportional to the major axis of the ellipse.*

The *minor axis* $2b$ of the ellipse, computed by setting $dy/dt = R(\varepsilon + \cos\theta) = 0$, is given by

$$b = y_{\max} = \Lambda/\sqrt{1-\varepsilon^2}.$$

The *area* of the ellipse is then given by

$$A = \pi ab = \pi\Lambda^2/(1-\varepsilon^2)^{3/2}.$$

Applying Kepler's Second Law in the form $dA/dt = h/2$, we see that the *period* T of such an elliptical orbit satisfies $A/T = h/2$. Together with (4), (5), and the definition of Λ, this implies that

$$T = 2\pi k/(-2E)^{3/2} = 2\pi(a^3/k)^{1/2}, \tag{6}$$

which is a modern statement of *Kepler's Third Law. Thus the period is proportional to* $a^{3/2}$ *and is inversely proportional to* $|E|^{3/2}$. For further details of these calculations see, for example, Arnold [1] or Synge and Griffith [21].

Remark. A familiar but incorrect version of this Third Law asserts that the period is proportional to the $3/2$ power of the "mean distance" of $\mathbf{x}(t)$ from the origin. In fact the mean distance, that is the time average $\oint r\,dt/\oint dt$, is not equal to the major semi-axis a but is rather equal to $(1+\varepsilon^2/2)a$. However it is interesting to note that the time average of r^{-1} is precisely equal to a^{-1}. Proofs based on Formula (9) of Appendix 1 are easily supplied.

It is noteworthy that *the period of such a periodic orbit depends only on the energy*. This is true for period solutions of Lagrangian or Hamiltonian differential equations under quite general conditions. Compare Herglotz [9], Wintner [22, §100], and Gordon [6].

2. The Levi-Civita Metric. Moser [16] has given a very pretty picture of the space of all elliptic orbits of fixed energy, or fixed period, in terms of *stereographic projection*. (See also Györgyi [7].) To carry out Moser's construction, we must introduce a 3-dimensional sphere, consisting of all unit vectors $\mathbf{u} = (u_1, u_2, u_3, u_4)$ in Euclidean 4-space, and project it stereographically from its

"north pole" $(0,0,0,1)$ onto the 3-dimensional Euclidean space consisting of vectors of the form $\mathbf{v} = (v_1, v_2, v_3, 0)$. Thus each unit vector $\mathbf{u}$ maps to the unique vector $\mathbf{v} = (u_1, u_2, u_3, 0)/(1 - u_4)$ which has last coordinate zero, and lies on the same straight line through the north pole. *Then every circle on the 3-sphere projects onto a circle or line in 3-space.* (Compare Appendix 2.)

A brief computation shows that the *antipodal map* $\mathbf{u} \mapsto -\mathbf{u}$ from the 3-sphere to itself corresponds to the *negative inversion map* $\mathbf{v} \mapsto -\mathbf{v}/|\mathbf{v}|^2$ from Euclidean 3-space to itself. In other words, diametrically opposite points of the 3-sphere correspond to points $\mathbf{v}_1$ and $\mathbf{v}_2$ in Euclidean 3-space which lie on a common line through the origin, and satisfy $\mathbf{v}_1 \cdot \mathbf{v}_2 = -1$. Evidently the images of *great circles* on the 3-sphere are just those circles, or lines through the origin, which map into themselves under this negative inversion operation. Using Lemma 1, we see that a circle in 3-space has this invariance property if and only if it is one of our "velocity circles" with energy $E = (\mathbf{c} \cdot \mathbf{c} - R^2)/2$ equal to $-1/2$, or equivalently with period T equal to $2\pi k$. *Thus stereographic projection carries great circles on the 3-sphere precisely onto the velocity circles associated with elliptic orbits of energy* $-1/2$.

There is one apparent defect in this picture. Namely those great circles which pass through the poles correspond to straight lines through the origin in velocity space. These do not seem to fit into our picture. However this apparent defect is actually a virtue in disguise. For it enables us to give a precise description of the limiting behavior of elliptic orbits of fixed period as the angular momentum h tends to zero. If $h \to 0$, keeping E fixed and negative, it is easy to see that the minor axis of the orbit ellipse tends to zero, so that the ellipse flattens out and tends towards a straight line segment of length $-k/E$ with one end at the origin. In order for the orbit to vary continuously as $h \to 0$, we must adopt the following.

REFLECTION CONVENTION. *If an orbit has angular momentum $h = 0$, and if $dr/dt < 0$ for some values of t so that $\mathbf{x}(t)$ falls freely along a straight line towards the origin, then $\mathbf{x}(t)$ is reflected back along this same line when it hits the origin.*

(Compare Devaney [**5**], McGehee [**14**].) In fact it is natural to adopt this same Reflection Convention whether the energy is positive, negative or zero.

By rotating our coordinates, we can put such a singular orbit in the form $t \mapsto (r(t), 0, 0)$, where $r(t) \geq 0$ satisfies the differential equation $(dr/dt)^2/2 = E + k/r$. As an example, if $E = 0$, it follows that r is proportional to $(t - t_0)^{2/3}$. If $E < 0$, then the solution curve $r = r(t)$ is a cycloid, periodic in t, and smooth except for a cusp wherever $r = 0$. (See Appendix 1.) Note in particular that the derivative dr/dt tends to minus infinity as r decreases to zero, and then jumps to plus infinity. *If $\mathbf{x}(t)$ bounces off the origin, then the velocity vector $\mathbf{v}(t)$, moving along a straight line, sweeps through the "point at infinity" in 3-space.* This velocity line corresponds, under stereographic projection, to a great circle which sweeps through the north pole of the unit 3-sphere.

Thus, if we adopt this reflection convention, then there is a precise one-to-one correspondence between orbits of energy $-1/2$ and great circles on the unit 3-sphere. The situation for other negative values of E is the same, except for a scale change in 3-space.

According to Osipov [**17**], [**18**], and Belbruno [**2**], [**3**], there is an analogous description for parabolic or hyperbolic orbits. First consider a parabolic orbit, with energy $E = 0$. Recall that the velocity vector $\mathbf{v}(t)$ associated with such an orbit moves along a circle C which passes through the origin. *Let us transform this problem by applying the inversion operation* $\mathbf{v} \mapsto \mathbf{w} = \mathbf{v}/|\mathbf{v}|^2$. Then a circle C through the origin corresponds to a straight line in the space of vectors $\mathbf{w}$. (See Appendix 2.) *In this way we obtain a precise one-to-one correspondence between orbits of energy zero and straight lines in the Euclidean space consisting of all inverted velocity vectors* $\mathbf{w}$. In this picture, the orbits which bounce off the origin correspond to straight lines through the origin in inverted velocity space.

Finally let us look at the positive energy case, say $E = +1/2$. Since $\mathbf{v} \cdot \mathbf{v} > 2E = 1$, it follows that the inverted velocity vector $\mathbf{w} = \mathbf{v}/|\mathbf{v}|^2$ varies over the unit ball $|\mathbf{w}| < 1$. Using Lemma 1, we see that the corresponding velocity circles C are invariant under inversion, and hence intersect the

boundary of the unit ball orthogonally. *Thus each orbit of energy* $+1/2$ *corresponds to a circle arc* $t \mapsto \mathbf{w}(t)$ *which spans the unit ball* $|\mathbf{w}| < 1$, *and intersects its boundary 2-sphere orthogonally.* Orbits which bounce off the origin correspond to diameters spanning this unit 3-ball.

It is natural to compare this picture with the *conformal unit ball model* for the "hyperbolic" non-Euclidean geometry of Lobachevsky. (See Appendix 2.) In this model, discovered by Beltrami and later by Poincaré, points of hyperbolic space correspond to points in the Euclidean unit ball, and hyperbolic straight lines correspond to circle arcs or diameters which span the unit ball, meeting its boundary orthogonally. *Thus the orbits of energy* $+1/2$ *correspond precisely to "straight lines" in this model for hyperbolic 3-space.* There is an analogous picture for any positive value of E.

In terms of Riemannian geometry, we can put these three different constructions into one common framework as follows. Levi-Civita pointed out that it is possible to simplify solutions to Equation (1) by introducing a *fictitious time parameter* $s = \int dt/r$ along any orbit, where $r = |\mathbf{x}|$. (Compare Appendix 1.) Using this parameter, we will prove the following.

THEOREM 2 (Osipov and Belbruno). *Fixing some constant energy* E, *consider the space* M_E *consisting of all velocity vectors* $\mathbf{v}$ *for which* $\mathbf{v} \cdot \mathbf{v} > 2E$, *together with a single improper point* $\mathbf{v} = \infty$. *This space possesses one and only one Riemannian metric* ds^2 *so that the arc-length parameter* $\int ds$ *along any velocity circle* $t \mapsto \mathbf{v}(t)$ *is precisely equal to the Levi-Civita parameter* $\int dt/|x(t)|$. *This metric is smooth and complete, with constant curvature* $-2E$, *and its geodesics are precisely the circles or lines* $t \mapsto \mathbf{v}(t)$ *associated with Kepler orbits.*

In other words, there is a unique way of defining the "length" of a curve Γ in this space M_E of all compatible velocity vectors so as to coincide with Levi-Civita's integral $\int_\Gamma dt/r$. If $K = -2E$ is positive, then M_E, with this definition of length, is isometric to a 3-sphere of radius $1/\sqrt{K}$ in Euclidean 4-space; and the great circles on this sphere correspond to velocity circles. If $K = 0$, then M_E is isometric to Euclidean space; and if $K < 0$, it is isometric to hyperbolic space, with the unit of distance chosen appropriately. In particular, in all three cases, M_E can be made into a smooth manifold even in a neighborhood of the special point $\mathbf{v} = \infty$. Geodesics which pass through this special point correspond to Kepler orbits which bounce off the origin.

Proof. The Newton equation $d\mathbf{v}/dt = -k\mathbf{x}/r^3$ implies that $|d\mathbf{v}/dt| = k/r^2$. Dividing this equation by the definition $ds = dt/r$, and recalling the definition of E, we obtain $|d\mathbf{v}/ds| = k/r = \mathbf{v} \cdot \mathbf{v}/2 - E$. Therefore $|ds| = 2|d\mathbf{v}|/(\mathbf{v} \cdot \mathbf{v} - 2E)$, or in other words

$$ds^2 = 4\, d\mathbf{v} \cdot d\mathbf{v}/(\mathbf{v} \cdot \mathbf{v} - 2E)^2, \tag{7}$$

Thus there is one and only one Riemannian metric on M_E *which satisfies our condition, and it is given by this formula* (7). To describe what happens in a neighborhood of infinity, we work with the inverted velocity coordinate $\mathbf{w} = \mathbf{v}/|\mathbf{v}|^2$. Note that $2E\mathbf{w} \cdot \mathbf{w} < 1$. Computation shows that $d\mathbf{w} \cdot d\mathbf{w} = d\mathbf{v} \cdot d\mathbf{v}/(\mathbf{v} \cdot \mathbf{v})^2$. (See Formula (10) of Appendix 2.) Hence $ds^2 = 4\, d\mathbf{w} \cdot d\mathbf{w}/(1 - 2E\mathbf{w} \cdot \mathbf{w})^2$, where the denominator is always strictly positive. If we set $K = -2E$, this becomes

$$ds^2 = 4\, d\mathbf{w} \cdot d\mathbf{w}/(1 + K\mathbf{w} \cdot w)^2, \tag{7$'$}$$

Except for a scale change, this is just the form given by Riemann, in his inaugural dissertation, for a metric of constant curvature K. In the case $K = 0$, this metric is obviously flat, with straight lines as geodesics. In the positive curvature case, we can compare it with the standard metric on a 3-sphere of radius $1/\sqrt{K}$. It is not difficult to check that these two metrics correspond isometrically, via stereographic projection from the 3-sphere to a 3-plane passing through its center; and that great circles correspond precisely to our (inverted) velocity circles. Further details of the proof, particularly in the case $K < 0$, may be found in Appendix 2. ∎

COROLLARY. *The functions* $\mathbf{x}(t)$ *and* $\mathbf{v}(t)$ *depend smoothly on the time* t *and the initial conditions* $\mathbf{x}(0)$ *and* $\mathbf{v}(0)$, *even in the neighborhood of an orbit which bounces off the origin, so long as the vectors* $\mathbf{x}(0)$ *and* $\mathbf{x}(t)$ *themselves are not zero.*

Here is an explicitly worked out example, to illustrate the behavior of these functions. Consider the family of parallel straight lines $s \mapsto \mathbf{w} = (-s, \alpha, \beta)/2$ in inverted velocity space, depending on two real parameters α and β. These correspond to the family of parabolic orbits

$$\mathbf{x} = k\left(\alpha^2 + \beta^2 - s^2, 2\alpha s, 2\beta s\right)/2.$$

Not only this orbital position $\mathbf{x}$, but also the distance

$$r = |\mathbf{x}| = k\left(\alpha^2 + \beta^2 + s^2\right)/2$$

and the time

$$t = \int_0 r\, ds = ks\left(\alpha^2 + \beta^2 + s^2/3\right)/2$$

can be expressed as smooth functions of the three parameters α, β, s. Using the implicit function theorem, we can solve for s, and hence $\mathbf{x}$, as functions of α, β, t. These functions are smooth except precisely at those points where $\partial t/\partial s = |\mathbf{x}|$ is zero.

Similarly, we can look at the six parameter family of orbits depending on the initial position vector $\mathbf{x}(0) \neq \mathbf{0}$ and the initial velocity vector $\mathbf{v}(0)$. Using the identity

$$r = 2k/(\mathbf{v} \cdot \mathbf{v} - 2E) = 2k\mathbf{w} \cdot \mathbf{w}/(1 - 2E\mathbf{w} \cdot \mathbf{w}),$$

where $E = \mathbf{v}(0) \cdot \mathbf{v}(0) - k/|\mathbf{x}(0)|$, we see that r and $t = \int_0 r\, ds$ can be expressed as smooth functions of the parameters $\mathbf{x}(0) \neq \mathbf{0}$, $\mathbf{v}(0)$ and s. In fact it is not difficult to verify the identity

$$\mathbf{x} = 4k(2(\mathbf{w} \cdot \mathbf{w}')\mathbf{w} - (\mathbf{w} \cdot \mathbf{w})\mathbf{w}')/(1 - 2E\mathbf{w} \cdot \mathbf{w}),$$

where $\mathbf{w}' = d\mathbf{w}/ds$, which shows that $\mathbf{x}$ can also be expressed as a smooth function of these parameters. Again, if we use t in place of s as parameter, then we lose the smoothness of the resulting functions only at those points where $\mathbf{x}$ is precisely equal to the zero vector. ■

Appendix 1. Conic sections.

This will be a brief review of some classical constructions. An *ellipse*, with foci $\mathbf{f}_1$ and $\mathbf{f}_2$, can be defined as the set of all vectors $\mathbf{x}$ in the Euclidean plane which satisfy the equation

$$|\mathbf{x} - \mathbf{f}_1| + |\mathbf{x} - \mathbf{f}_2| = 2a.$$

Here $2a > |\mathbf{f}_1 - \mathbf{f}_2|$ is a constant equal to the *major axis* of the ellipse. The ratio $\varepsilon = |\mathbf{f}_1 - \mathbf{f}_2|/2a$

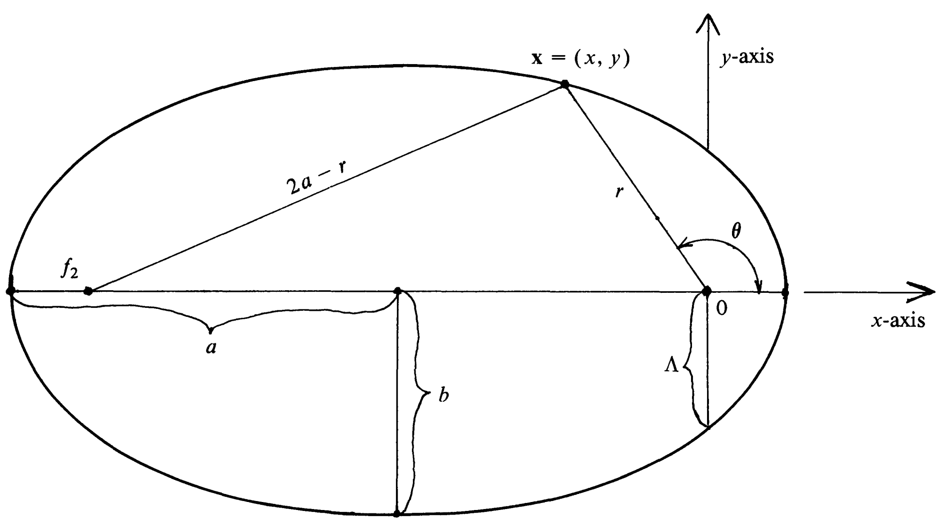


FIG. 2

of the distance between foci to the major axis is called the *eccentricity* of the ellipse. Note that $0 \le \varepsilon < 1$, where $\varepsilon = 0$ only for a circle. If we choose cartesian coordinates x, y so that $\mathbf{f}_1$ is the origin and $\mathbf{f}_2$ is the point $(-2\varepsilon a, 0)$, then the distances r and $2a - r$ of (x, y) from $\mathbf{f}_1$ and $\mathbf{f}_2$ are given by the equations

$$x^2 + y^2 = r^2, \quad (x + 2\varepsilon a)^2 + y^2 = (2a - r)^2.$$

(Fig. 2.) Subtracting one equation from the other and dividing by $4a$, we obtain

$$r = \sqrt{x^2 + y^2} = \Lambda - \varepsilon x, \tag{8}$$

where Λ denotes the constant $(1 - \varepsilon^2)a$, known as the "semi-latus-rectum."

More generally, if we fix any constants $\varepsilon \ge 0$ and $\Lambda > 0$, then the locus described by Equation (8) is called a *conic section* with focus at the origin, and with eccentricity ε. Squaring both sides of (8), we obtain

$$(1 - \varepsilon^2)x^2 + y^2 = \Lambda^2 - 2\Lambda\varepsilon x.$$

Evidently this equation describes a *parabola* if $\varepsilon = 1$, or a *hyperbola* if $\varepsilon > 1$. Substituting $x = r\cos\theta$, $y = r\sin\theta$ in (8), we obtain the equation

$$r = \Lambda/(1 + \varepsilon\cos\theta) \tag{3}$$

which defines such a conic section, with focus at the origin, in polar coordinates. In the elliptic case $\varepsilon < 1$, note that the quadratic equation relating x and y can be put in the form $(x + \varepsilon a)^2/a^2 + y^2/b^2 = 1$, where $b = a\sqrt{1 - \varepsilon^2}$. Thus this ellipse is centered at the point $(-\varepsilon a, 0)$, and has principal axes $2a \ge 2b$.

Given the constant $k > 0$, we can introduce the time t into this geometrical picture by means of Kepler's Second Law, $d\theta/dt = h/r^2$, or in other words $t = \int r^2\, d\theta/h$, where $h = \sqrt{k\Lambda}$. Similarly, Levi-Civita's fictitious time can be introduced as the integral $s = \int dt/r = \int r\, d\theta/h$ along an orbit. In practice, it turns out to be easiest to express both time and position as functions of s. Substituting (3) in the equations $x = r\cos\theta$, $y = r\sin\theta$ and differentiating, we find that

$$dx/ds = -Ry \quad dy/ds = h\varepsilon + (1 - \varepsilon^2)Rx,$$

and therefore $d^2y/ds^2 = 2Ey$, where $R = h/\Lambda$, and $2E = -(1 - \varepsilon^2)R^2$. From this, it is not difficult to compute y, x, and $t = \int(\Lambda - \varepsilon x)\, ds$ as functions of s. As an example, in the negative energy case $2E = -\alpha^2 < 0$, if we normalize so that $s = 0$ at the point of closest approach, we find that

$$x = -\varepsilon a + a\cos\alpha s, \quad y = b\sin\alpha s,$$

with

$$r = a(1 - \varepsilon\cos\alpha s), \quad t = a(s - \varepsilon\alpha^{-1}\sin\alpha s). \tag{9}$$

(Compare KEPLER.) This form of the equations again makes the elliptical shape of the orbit quite clear. In Kepler's terminology, the angle αs is known as the "eccentric anomaly," and the angle $\alpha t/a = 2\pi t/T$ is called the "mean anomaly." If the eccentricity ε tends to 1, keeping α and $a = k/\alpha^2$ fixed, so that b and Λ tend to zero, note that these expressions tend to well behaved limits. The resulting curve in the t, x plane is known as a *cycloid*. In the positive energy case, there are completely analogous formulas involving the hyperbolic sine and cosine functions. In the zero energy case, evidently y is a linear function of s, hence x is a quadratic function, and t is a cubic function of s.

Appendix 2. Inversive geometry, stereographic projection, and hyperbolic geometry.

This will be a quick review of well-known material. Further details may be found for example in Coxeter [4] or in Hilbert and Cohn-Vossen [**10**].

Let **p** be a base point in the n-dimensional Euclidean space E. The operation of *inversion*, in the unit sphere centered at **p**, is a smooth mapping from $E - \mathbf{p}$ to itself, defined as follows. *Each point* $\mathbf{x} \neq \mathbf{p}$ *of Euclidean space maps to the unique point* **y** *which lies on the ray* (= *half-line*) *which starts at* **p** *and passes through* **x**, *such that the Euclidean distance* $|\mathbf{y} - \mathbf{p}|$ *is equal to the reciprocal* $1/|\mathbf{x} - \mathbf{p}|$. Thus each point of the unit sphere $|\mathbf{x} - \mathbf{p}| = 1$ is fixed by this inversion map, but the inside and outside of the sphere are interchanged. More generally, given any constant $r > 0$, we can invert in the sphere of radius r centered at **p**. The definition is the same, except that we set $|\mathbf{y} - \mathbf{p}|$ equal to $r^2/|\mathbf{x} - \mathbf{p}|$.

It is often convenient to extend this construction by adjoining a formal *point at infinity* to Euclidean space. Then inversion maps $E \cup \infty$ to itself, with the understanding that the base point **p** maps to the point ∞ and that ∞ maps to **p**.

Here is an example to illustrate the inversion map. *Consider an* $(n-1)$-*dimensional sphere* S *in Euclidean space, with two diametrically opposite points* **p** *and* **q**, *and let* P *be an* $(n-1)$-*dimensional hyperplane, situated as in Figure* 3, *so that the half-line from* **p** *through* **q** *meets* P *orthogonally at a point* $\mathbf{z} \neq \mathbf{p}$. Then for any **x** of S and **y** of P which lie on a common line through **p**, the right triangle **p**, **x**, **q** is similar to the right triangle **p**, **z**, **y**. Therefore the ratio $|\mathbf{x} - \mathbf{p}|/|\mathbf{q} - \mathbf{p}|$ is equal to the ratio $|\mathbf{z} - \mathbf{p}|/|\mathbf{y} - \mathbf{p}|$. In other words, the distance from **x** to **p** is inversely proportional to the distance from y to p. *This proves that the sphere* S *maps onto the hyperplane* $P \cup \infty$ *under inversion from* **p**, *provided that the constant* r *is chosen correctly*; *and conversely it proves that* $P \cup \infty$ *maps onto* S.

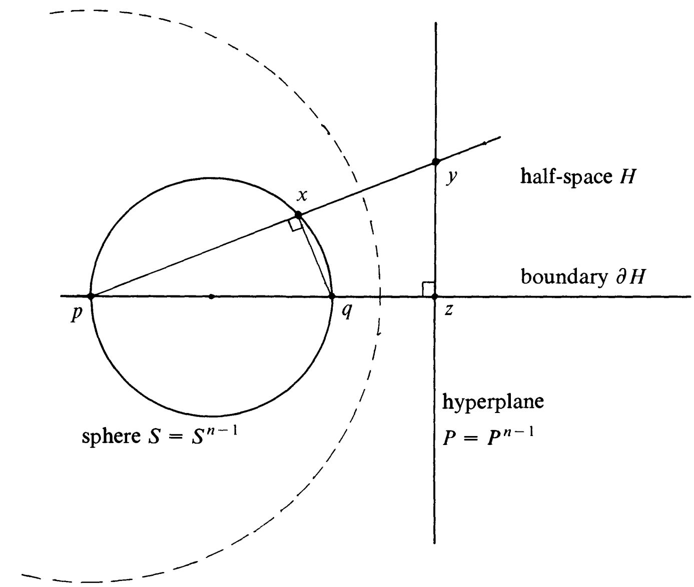


FIG. 3

Definition. The one-to-one correspondence between the sphere S and the hyperplane $P \cup \infty$ which is constructed in this way is called *stereographic projection* from **p**. Geometrically, stereographic projection maps each point $\mathbf{x} \neq \mathbf{p}$ of S to the unique point **y** of P such that **x**, **y**, and **p** all lie on a common straight line.

If P^k is a plane of dimension k in Euclidean n-space, where k can be any number between 1 and $n-1$, then a completely analogous argument shows that the image of $P^k \cup \infty$ under inversion is either a sphere S^k (if P^k does not contain the base point **p**), or is equal to $P^k \cup \infty$ itself if P^k does contain **p**. More generally, we will prove the following.

LEMMA 2 (Steiner, 1824). *The operation of inversion carries every straight line or circle in $E \cup \infty$ to a straight line or circle. Similarly it carries every k-dimensional sphere or plane to a k-dimensional sphere or plane. Furthermore it preserves angles. That is, if two lines or circles intersect at angle θ (at a point other than* $\mathbf{p}$), *then their images under inversion also intersect at angle θ.*

To make sense of this statement, we must adopt the convention that the point at infinity belongs to every straight line or hyperplane, but that it does not belong to any circle.

As an immediate corollary of this lemma: *Stereographic projection maps every circle in the sphere S to a circle or straight line in the hyperplane $P \cup \infty$. Furthermore, stereographic projection also preserves angles.*

Proof of Lemma 2. First consider a sphere of dimension $n - 1$. By the remarks above, we need only consider the case where this sphere S^{n-1} does not contain the base point $\mathbf{p}$. It will be convenient to choose cartesian coordinates so that $\mathbf{p}$ is the origin. If $\mathbf{x}$ and $\mathbf{x}'$ are two points of S^{n-1} which lie on a common line through the origin $\mathbf{p}$, then, as in the proof of Lemma 1, the scalar product $\mathbf{x} \cdot \mathbf{x}'$ is equal to a nonzero constant, which we can write as r^2/a. It follows that the image of $\mathbf{x}'$ under inversion is equal to $a\mathbf{x}$. *Therefore, the image of S^{n-1} under inversion is precisely equal to the sphere consisting of all points $a\mathbf{x}$ with $\mathbf{x}$ in S^{n-1}.*

The proof for a sphere S^k of arbitrary dimension now follows easily. For S^k can be described as the intersection of the sphere S^{n-1} which has the same center and the same radius with a plane P^{k+1}. Since inversion maps both S^{n-1} and P^{k+1} to a sphere or plane of the same dimension, it follows that it maps their intersection to a sphere or plane.

Next consider two smooth curves Γ_1 and Γ_2 in Euclidean space which intersect at a point $\mathbf{x} \neq \mathbf{p}$. Then their images $\Gamma_1{}^*$ and $\Gamma_2{}^*$ under inversion intersect at a corresponding point $\mathbf{x}^*$. If θ is the angle between the tangent vectors of the two curves at $\mathbf{x}$, and θ^* is the angle between the corresponding tangent vectors at $\mathbf{x}^*$, we must prove that $\theta = \theta^*$. Let L_i be the straight line which is tangent to Γ_i at $\mathbf{x}$, let L_i' be the parallel straight line through $\mathbf{p}$, and let C_i be the image of L_i under inversion. Then L_i' is tangent to C_i at $\mathbf{p}$, so the angle between C_1 and C_2 at $\mathbf{p}$ is equal to θ. Evidently the angle between these two circles at their two points of intersection must be equal. Since C_i is tangent to $\Gamma_i{}^*$ at $\mathbf{x}^*$, this completes the proof. ∎

Now let us describe a very different kind of geometry. *Non-Euclidean geometry*, also called *hyperbolic geometry*, was first introduced by Lobachevsky and Bolyai by means of a collection of geometric axioms. These were identical to the axioms of Euclidean geometry, except for the hypothesis that, within a plane, it is possible to construct more than one parallel to a given line through a given point. There was no proof that these axioms were consistent for many years, until Beltrami showed that hyperbolic geometry could be modeled within Euclidean geometry. (See Milnor [**15**] for references.) One of Beltrami's Euclidean realizations of hyperbolic geometry was the *upper half-space model*, later utilized by Poincaré. It can be described as follows.

Let H be an open half-space, with boundary ∂H, in the n-dimensional Euclidean space E. It will be convenient to make use of Cartesian coordinates $\mathbf{x} = (x_1, \cdots, x_n)$, chosen so that the half-space H is defined by the inequality $x_n > 0$. Thus ∂H is the hyperplane $x_n = 0$. If Γ is a smooth curve, described parametrically by the equation $\mathbf{x} = \mathbf{x}(t)$, then we will use the notation $\int_\Gamma |d\mathbf{x}|$ for the Euclidean length $\int |d\mathbf{x}/dt|\, dt$.

Definition. By the *hyperbolic length* of a smooth curve Γ in the half-space H will be meant the integral $\int_\Gamma |d\mathbf{x}|/x_n$, along Γ, of the Euclidean length element $|d\mathbf{x}|$ divided by the Euclidean distance from ∂H.

A curve Γ in H will be called a *hyperbolic line* if it provides the hyperbolically-shortest possible path between any two of its points. In other words, there must be no curve in H which joins two points of Γ and has hyperbolic length strictly less than the hyperbolic length of the segment Γ_0 of Γ between these points. (In Riemannian geometry, such a curve Γ is called a *minimal geodesic*.)

LEMMA 3 (Beltrami, 1868). *A curve in the half-space H is a hyperbolic line if and only if it is*

either a Euclidean half-line which meets the boundary of H orthogonally, or a Euclidean semi-circle which meets ∂H orthogonally.

Proof. First suppose that Γ is a half-line

$$x_1 = \text{constant}, \ldots, x_{n-1} = \text{constant}, x_n > 0.$$

If Δ is any other curve segment joining two points of Γ, then it is easy to check that

$$\int_\Delta |d\mathbf{x}|/x_n \geq \int_\Delta |dx_n|/x_n \geq \int_{\Gamma_0} |dx_n|/x_n.$$

In fact there is a strict inequality unless Δ is also a vertical line segment. *This proves that Γ is a hyperbolic line; and furthermore that it provides the unique hyperbolically-shortest path between any two of its points.*

Now consider an inversion mapping $\mathbf{x} \mapsto \mathbf{y}$, using any base point $\mathbf{p}$ on ∂H. To simplify the computation, let us first suppose that $\mathbf{p} = \mathbf{0}$. Differentiating the equation $\mathbf{y} = r^2\mathbf{x}/\mathbf{x}\cdot\mathbf{x}$, we obtain

$$d\mathbf{y} = r^2((\mathbf{x}\cdot\mathbf{x})\,d\mathbf{x} - 2(\mathbf{x}\cdot d\mathbf{x})\mathbf{x})/(\mathbf{x}\cdot\mathbf{x})^2.$$

If we take the dot product of this equation with itself, the terms involving $\mathbf{x}\cdot d\mathbf{x}$ cancel, so that we obtain

(10) $$d\mathbf{y}\cdot d\mathbf{y} = r^4\,d\mathbf{x}\cdot d\mathbf{x}/(\mathbf{x}\cdot\mathbf{x})^2.$$

More generally, for any choice of base point p, the appropriate formula is

$$d\mathbf{y}\cdot d\mathbf{y} = d\mathbf{x}\cdot d\mathbf{x}(r/|\mathbf{x}-\mathbf{p}|)^4.$$

It follows easily that

$$|d\mathbf{y}|/|\mathbf{y}-\mathbf{p}| = |d\mathbf{x}|/|\mathbf{x}-\mathbf{p}|.$$

since the vector $\mathbf{y}-\mathbf{p}$ is a positive multiple of $\mathbf{x}-\mathbf{p}$, this implies that

$$|d\mathbf{y}|/y_n = |d\mathbf{x}|/x_n,$$

whenever the base point p belongs to the boundary of H. *Thus the inversion mapping* $\mathbf{x} \mapsto \mathbf{y}$, *with any choice of base point in ∂H, maps the half-space H into itself so as to preserve hyperbolic length.* Such a length preserving mapping from H to itself is called a *hyperbolic isometry*.

It follows from Lemma 2 that inversion carries any half-line meeting ∂H orthogonally to a semicircle meeting ∂H orthogonally. Since we have shown that these Euclidean half-lines are hyperbolic lines, it follows that *these semi-circles must also be hyperbolic lines.*

To see that there are no other hyperbolic lines, suppose that we start with two arbitrary points $\mathbf{x}$ and $\mathbf{y}$ of the half-space H. Then it is not difficult to check that $\mathbf{x}$ and $\mathbf{y}$ lie on either a Euclidean half-line or a Euclidean semi-circle which meets ∂H orthogonally. Thus we have constructed enough hyperbolic lines to join any two points of H; and it follows that we have constructed *all* hyperbolic lines. ■

More generally, a k-dimensional subset of H is called a *hyperbolic k-plane* if it contains the hyperbolic line joining any two of its points. A similar argument shows that a subset of H is a hyperbolic k-plane if and only if it is either a Euclidean half-plane or a Euclidean hemisphere, meeting the boundary of H orthogonally, and having dimension k.

Note that the group consisting of all hyperbolic isometries from H to itself is quite large. The proof above shows that any inversion map, with base point in ∂H, belongs to this group of isometries. As a further example, it is not difficult to check that the mapping

$$(x_1,\cdots,x_n) \mapsto c_n(x_1,\cdots,x_n) + (c_1,\cdots,c_{n-1},0)$$

is a hyperbolic isometry which carries the point $\mathbf{u} = (0, \cdots, 0, 1)$ to a completely arbitrary point $\mathbf{c} = (c_1, \cdots, c_n)$ of H.

One characteristic property of this geometry is the following. *In any hyperbolic plane, consider a triangle Δ of hyperbolic area A, bounded by three hyperbolic line segments. The sum of the interior angles of Δ is equal to $\pi - A$.* Here we use the usual Euclidean definition of angle in the upper half-space model, but A must be defined as the integral $\int_\Delta (dA)_{\text{Euclidean}}/(x_n)^2$ so as to be invariant under hyperbolic isometries; where $(dA)_{\text{Euclidean}}$ stands for the usual Euclidean area element.

Proof. We may assume that Δ lies in the plane $x_2 = \cdots = x_{n-1} = 0$. Setting $x = x_1, y = x_n$, and integrating $A = \int\int_\Delta dx\,dy/y^2$ with respect to the y variable, we obtain $A = \int_{\partial\Delta} dx/y$, to be integrated around the boundary of Δ. Each edge of Δ is either a vertical line segment, on which dx/y is zero, or a circle arc

$$x = x_0 + a\cos\theta, \quad y = a\sin\theta,$$

on which $dx/y = -d\theta$. Thus $A = -\int_{\partial\Delta} d\theta$. An elementary argument then shows that this expression is equal to π minus the sum of the angles. ∎

By way of contrast, for any triangle on the unit 2-sphere the sum of the interior angles is equal to π *plus* the area. (Compare Coxeter [**4**], pp. 95, 297.) *On a sphere of radius r, the corresponding formula would be*

$$\Sigma(\text{interior angles}) = \pi + KA, \tag{11}$$

where $K = 1/r^2$. Let us take this last formula as a definition of the *curvature* of a space of constant curvature K. Then evidently Euclidean space has curvature $K = 0$, and hyperbolic space has curvature $K = -1$. More generally, if we change the definition of "length," multiplying all hyperbolic lengths by r and all areas by r^2, then we obtain a space of constant curvature $K = -1/r^2$.

Section 2 makes use of the *conformal unit ball model* for hyperbolic space. This is a slightly different Euclidean model which can be constructed as follows. Let us invert the upper half-space H, consisting of all points $\mathbf{x}$ with $x_n > 0$, with respect to a sphere of radius $r = \sqrt{2}$ centered at the point $\mathbf{p} = (0, \ldots, 0, -1)$. Then it is not difficult to check that the image of H under this inversion is precisely the open unit ball B, consisting of all vectors $\mathbf{y}$ with $|\mathbf{y}| < 1$. Let $\mathbf{u}$ be the upward unit vector $-\mathbf{p} = (0, \cdots, 0, 1)$. A brief computation, based on the formulas $\mathbf{x} + \mathbf{u} = 2(\mathbf{y} + \mathbf{u})/|\mathbf{y} + \mathbf{u}|^2$ and $|d\mathbf{x}| = 2|d\mathbf{y}|/|\mathbf{y} + \mathbf{u}|^2$, shows that the hyperbolic length element $|d\mathbf{x}|/x_n = |d\mathbf{x}|/\mathbf{x} \cdot \mathbf{u}$ in the half-space H corresponds to the length element

$$2|d\mathbf{y}|/(1 - \mathbf{y} \cdot \mathbf{y}) \tag{12}$$

in the ball B. *This last expression is the Riemann-Beltrami formula for the hyperbolic length element in the unit ball* $\mathbf{y} \cdot \mathbf{y} < 1$.

We can take this ball B, with the length element (12), as an alternative model for hyperbolic space. It follows from Lemmas 2 and 3 that the *hyperbolic lines* in this model are just the circle arcs or diameters which span the ball B, meeting its boundary sphere orthogonally.

This model gives us a further understanding of the richness of the group of isometries of hyperbolic space, for the subgroup consisting of all hyperbolic isometries of B which fix the origin can evidently be identified with the orthogonal group, consisting of all isometries of Euclidean space fixing the origin. The existence of such a large group of isometries would provide a key step in a proof that hyperbolic geometry, constructed in this way, satisfies all of the axioms of Euclidean geometry with the exception of the parallel axiom.

References

1. V. I. Arnold, Mathematical Methods of Classical Mechanics, Graduate Texts in Math, 60, Springer, 1978.
2. E. A. Belbruno, Two body motion under the inverse square central force and equivalent geodesic flows, Celest. Mech., 15 (1977) 467–476.

3. ______, Regularizations and geodesic flows, pp. 1–11 of Classical Mechanics and Dynamical Systems, R. L. Devaney and Z. H. Nitecki, Editors, Dekker, 1981.

4. H. S. M. Coxeter, Introduction to Geometry, 2nd ed., Wiley, 1969.

5. R. L. Devaney, Blowing up singularities in classical mechanical systems, this MONTHLY, 89 (1982) 535–552.

6. W. B. Gordon, On the relation between period and energy in periodic dynamical systems, J. Math. Mech., 19 (1969) 111–114.

7. G. Györgyi, Kepler's equation, Fock variables, Bacry's generators and Dirac brackets, Nuovo Cimento 53A, (1968) 717–736.

8. W. R. Hamilton, The hodograph or a new method of expressing in symbolic language the Newtonian law of attraction, Proc. Roy. Irish Acad., 3 (1846) 344–353 (Math. Papers v. 2, 287–294, Camb. U. Press, 1940).

9. G. Herglotz, Bemerkungen zum dritten Keplerschen Gesetz, Probleme der Astronomie: Festschrift für Hugo v. Seeliger, Springer Berlin, 1924, 197–199.

10. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, 1952.

11. J. Kepler, Astronomia Nova, Prague 1609, §59,60 (cf. "Neue Astronomie," Munich-Berlin 1929, 412–413, or Gesam, Werke 3, Munich, 1937, 480–482).

12. M. Kummer, On the regularization of the Kepler problem, Comm. Math. Phys., 84 (1982) 133–152.

13. T. Levi-Civita, Fragen der klassischen und relativistischen Mechanik, Springer, 1924.

14. R. McGehee, Singularities in classical celestial mechanics, Proc. Int. Congr. Math., Helsinki, 1978, 827–834.

15. J. Milnor, Hyperbolic geometry: the first 150 years, Bull. Amer. Math. Soc., 6 (1982) 9–24.

16. J. Moser, Regularization of Kepler's problem and the averaging method on a manifold, Comm. Pure App. Math., 23 (1970) 609–636.

17. Yu. S. Osipov, Geometrical interpretation of Kepler's problem (Russian), Uspehi Mat. Nauk, 27 #2 (1972) p. 161.

18. ______, The Kepler problem and geodesic flows in spaces of constant curvature, Celest. Mech., 16 (1977) 191–208.

19. B. Riemann, Ueber die Hypothesen welche der Geometrie zu Grunde liegen, Abh. K. G. Wiss. Göttingen, 13 (1868).

20. E. L. Stiefel and G. Scheifele, Linear and Regular Celestial Mechanics, Springer, 1971.

21. J. L. Synge and B. A. Griffith, Principles of Mechanics, McGraw Hill, 2nd ed., 1949.

22. A. Wintner, The Analytical Foundations of Celestial Mechanics, Princeton U. Press, 1941.

SHORT THEOREMS WITH LONG PROOFS

JOEL SPENCER

Department of Mathematics, State University of New York, Stony Brook, NY 11794

Long proofs are an anathema to mathematicians. Part of mathematics' uniqueness is the verifiability of an argument. This is being called into question with the existence of proofs of inordinate length. A folk theorem, surely known in the 1930's, is particularly germane today. A recent note in which F. H. Norwood [1] appears unaware of this result provided the motivation for these remarks. We give three variants, with increasing formality.

(i) There exist short theorems with long proofs.

(ii) There exist theorems T whose shortest proofs have length at least $10^{100}2^{2^n}$ where n is the length of T.

(iii) For all recursive functions F there exist theorems T whose shortest proofs have length at least $F(n)$ where n is the length of T.

Joel Spencer received his Ph.D. degree from Harvard University in 1970 under the direction of Andrew Gleason. His research interests are in combinatorial analysis, where he considers himself a disciple of Paul Erdős. He has taught at UCLA, MIT, and, since 1975, at SUNY at Stony Brook. He spent the academic years 1976-77 in Budapest and 1980-81 in Rehovot (Israel) and Reading (England). He enjoys coaching Putnam teams for his students and soccer teams for his children.

3. ______, Regularizations and geodesic flows, pp. 1–11 of Classical Mechanics and Dynamical Systems, R. L. Devaney and Z. H. Nitecki, Editors, Dekker, 1981.

4. H. S. M. Coxeter, Introduction to Geometry, 2nd ed., Wiley, 1969.

5. R. L. Devaney, Blowing up singularities in classical mechanical systems, this MONTHLY, 89 (1982) 535–552.

6. W. B. Gordon, On the relation between period and energy in periodic dynamical systems, J. Math. Mech., 19 (1969) 111–114.

7. G. Györgyi, Kepler's equation, Fock variables, Bacry's generators and Dirac brackets, Nuovo Cimento 53A, (1968) 717–736.

8. W. R. Hamilton, The hodograph or a new method of expressing in symbolic language the Newtonian law of attraction, Proc. Roy. Irish Acad., 3 (1846) 344–353 (Math. Papers v. 2, 287–294, Camb. U. Press, 1940).

9. G. Herglotz, Bemerkungen zum dritten Keplerschen Gesetz, Probleme der Astronomie: Festschrift für Hugo v. Seeliger, Springer Berlin, 1924, 197–199.

10. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, 1952.

11. J. Kepler, Astronomia Nova, Prague 1609, §59,60 (cf. "Neue Astronomie," Munich-Berlin 1929, 412–413, or Gesam, Werke 3, Munich, 1937, 480–482).

12. M. Kummer, On the regularization of the Kepler problem, Comm. Math. Phys., 84 (1982) 133–152.

13. T. Levi-Civita, Fragen der klassischen und relativistischen Mechanik, Springer, 1924.

14. R. McGehee, Singularities in classical celestial mechanics, Proc. Int. Congr. Math., Helsinki, 1978, 827–834.

15. J. Milnor, Hyperbolic geometry: the first 150 years, Bull. Amer. Math. Soc., 6 (1982) 9–24.

16. J. Moser, Regularization of Kepler's problem and the averaging method on a manifold, Comm. Pure App. Math., 23 (1970) 609–636.

17. Yu. S. Osipov, Geometrical interpretation of Kepler's problem (Russian), Uspehi Mat. Nauk, 27 #2 (1972) p. 161.

18. ______, The Kepler problem and geodesic flows in spaces of constant curvature, Celest. Mech., 16 (1977) 191–208.

19. B. Riemann, Ueber die Hypothesen welche der Geometrie zu Grunde liegen, Abh. K. G. Wiss. Göttingen, 13 (1868).

20. E. L. Stiefel and G. Scheifele, Linear and Regular Celestial Mechanics, Springer, 1971.

21. J. L. Synge and B. A. Griffith, Principles of Mechanics, McGraw Hill, 2nd ed., 1949.

22. A. Wintner, The Analytical Foundations of Celestial Mechanics, Princeton U. Press, 1941.

SHORT THEOREMS WITH LONG PROOFS

JOEL SPENCER

Department of Mathematics, State University of New York, Stony Brook, NY 11794

Long proofs are an anathema to mathematicians. Part of mathematics' uniqueness is the verifiability of an argument. This is being called into question with the existence of proofs of inordinate length. A folk theorem, surely known in the 1930's, is particularly germane today. A recent note in which F. H. Norwood [**1**] appears unaware of this result provided the motivation for these remarks. We give three variants, with increasing formality.

(i) There exist short theorems with long proofs.

(ii) There exist theorems T whose shortest proofs have length at least $10^{100}2^{2^n}$ where n is the length of T.

(iii) For all recursive functions F there exist theorems T whose shortest proofs have length at least $F(n)$ where n is the length of T.

Joel Spencer received his Ph.D. degree from Harvard University in 1970 under the direction of Andrew Gleason. His research interests are in combinatorial analysis, where he considers himself a disciple of Paul Erdős. He has taught at UCLA, MIT, and, since 1975, at SUNY at Stony Brook. He spent the academic years 1976-77 in Budapest and 1980-81 in Rehovot (Israel) and Reading (England). He enjoys coaching Putnam teams for his students and soccer teams for his children.

There remains a hidden variable: the formal system in which theorems are proven. The folk theorem applies to any formal system for which there is no decision procedure for theoremhood. This includes Peano Arithmetic, Zermelo-Fraenkel set theory, or any formalization of "Generally Accepted Mathematics." It does not apply to limited systems such as the first order theory of Finite Fields.

Suppose (iii) were false for a specific F and every theorem of length n had a proof of length at most $F(n)$. This would give a simple decision procedure for statements S. Let n be the length of S and examine all texts of length at most $F(n)$. The statement S is a theorem if and only if one of these texts is a proof of S. As no such decidability procedure exists, (iii) is proved.

Version (i), even allowing for informality, is somewhat misleading. Since the integer n which exists by the folk theorem may itself be very large, it may be the case that there are long theorems with very, very long proofs.

Einstein's dogma, "God does not play dice with the universe" may be translated from physics to mathematics in (at least) two ways:

(a) Short interesting statements are decidable.

(b) Short interesting theorems have short proofs.

Fifty years after Gödel, many mathematicians view (a) as false. (The Axiom of Choice and the Continuum Hypothesis are short, interesting, and known to be undecidable, but let us restrict ourselves to statements expressible in Peano Arithmetic.) "Are there an infinite number of primes of the form $2^p - 1$?" may well be undecidable. The advent of computer proofs has focused attention on (b). A decade ago no counterexamples to (b) were known. Now there are two. The Four-Color Theorem is the first. The cataloging of Finite Simple Groups (in 5000 pages) is the second—or, rather, a second class. Short interesting theorems such as "Every 6-transitive subgroup of S_n is $(n-2)$-transitive" now have proofs of $5000 + \varepsilon$ pages. In both cases shorter proofs are being sought and may eventually be found. Or perhaps not. The possible falsity of (b) is only now seeping into our mathematical consciousness.

Reference

1. F. H. Norwood, Long proofs, this MONTHLY, 89 (1982) 110-112.

FIXED-ROUTE COST ALLOCATION

P. C. FISHBURN AND H. O. POLLAK
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

1. Introduction. Not long ago one of us (H. P.) flew from Newark, New Jersey, to Columbus, Ohio, to Raleigh/Durham to Tampa to Washington, D.C., then back to Newark. He went to Columbus on Bell Laboratories business, to Raleigh/Durham for the State of North Carolina, to Tampa for personal reasons, and to Washington on National Science Foundation business. After returning home, he had to decide how much of his air fare to charge each sponsor of his trip, namely Bell Labs, the State of North Carolina, himself, and the NSF. He wished to do this in as fair a manner as possible. At the minimum, he felt that the sponsors should be charged exactly the total air fare, and that no sponsor should pay more than the direct round-trip cost between Newark and the sponsor's city.

Peter Fishburn joined the Mathematics and Statistics Research Center at Bell Labs in 1978 following apprenticeships at the Institute for Advanced Study, The Pennsylvania State University, and elsewhere. Henry Pollak joined Bell Labs in 1951 and has been the director of the M & SR Center since 1962. He was president of the MAA in 1975-6 and has served the Association in many other capacities.

There remains a hidden variable: the formal system in which theorems are proven. The folk theorem applies to any formal system for which there is no decision procedure for theoremhood. This includes Peano Arithmetic, Zermelo-Fraenkel set theory, or any formalization of "Generally Accepted Mathematics." It does not apply to limited systems such as the first order theory of Finite Fields.

Suppose (iii) were false for a specific F and every theorem of length n had a proof of length at most $F(n)$. This would give a simple decision procedure for statements S. Let n be the length of S and examine all texts of length at most $F(n)$. The statement S is a theorem if and only if one of these texts is a proof of S. As no such decidability procedure exists, (iii) is proved.

Version (i), even allowing for informality, is somewhat misleading. Since the integer n which exists by the folk theorem may itself be very large, it may be the case that there are long theorems with very, very long proofs.

Einstein's dogma, "God does not play dice with the universe" may be translated from physics to mathematics in (at least) two ways:

(a) Short interesting statements are decidable.

(b) Short interesting theorems have short proofs.

Fifty years after Gödel, many mathematicians view (a) as false. (The Axiom of Choice and the Continuum Hypothesis are short, interesting, and known to be undecidable, but let us restrict ourselves to statements expressible in Peano Arithmetic.) "Are there an infinite number of primes of the form $2^p - 1$?" may well be undecidable. The advent of computer proofs has focused attention on (b). A decade ago no counterexamples to (b) were known. Now there are two. The Four-Color Theorem is the first. The cataloging of Finite Simple Groups (in 5000 pages) is the second—or, rather, a second class. Short interesting theorems such as "Every 6-transitive subgroup of S_n is $(n-2)$-transitive" now have proofs of $5000 + \varepsilon$ pages. In both cases shorter proofs are being sought and may eventually be found. Or perhaps not. The possible falsity of (b) is only now seeping into our mathematical consciousness.

Reference

1. F. H. Norwood, Long proofs, this MONTHLY, 89 (1982) 110-112.

FIXED-ROUTE COST ALLOCATION

P. C. FISHBURN AND H. O. POLLAK
Bell Telephone Laboratories, Murray Hill, New Jersey 07974

1. Introduction. Not long ago one of us (H. P.) flew from Newark, New Jersey, to Columbus, Ohio, to Raleigh/Durham to Tampa to Washington, D.C., then back to Newark. He went to Columbus on Bell Laboratories business, to Raleigh/Durham for the State of North Carolina, to Tampa for personal reasons, and to Washington on National Science Foundation business. After returning home, he had to decide how much of his air fare to charge each sponsor of his trip, namely Bell Labs, the State of North Carolina, himself, and the NSF. He wished to do this in as fair a manner as possible. At the minimum, he felt that the sponsors should be charged exactly the total air fare, and that no sponsor should pay more than the direct round-trip cost between Newark and the sponsor's city.

Peter Fishburn joined the Mathematics and Statistics Research Center at Bell Labs in 1978 following apprenticeships at the Institute for Advanced Study, The Pennsylvania State University, and elsewhere. Henry Pollak joined Bell Labs in 1951 and has been the director of the M & SR Center since 1962. He was president of the MAA in 1975-6 and has served the Association in many other capacities.

Our purpose here is to consider allocation schemes based on fairness criteria for situations similar to the one just described. We presume that each sponsor's willingness to pay for intercity travel reflects a component of the benefit the sponsor anticipates from the visit. Other benefit measures will not be considered explicitly at this time, but when they are available we would like to incorporate them in the analysis. Moreover, until the final section, willingness to pay will be assumed to equal the direct round-trip cost between the home city and the sponsor's city.

An allocation scheme that seems sensible to many people on first impression is a marginal-cost procedure that charges each sponsor at least the incremental cost caused by his participation or, if these increments sum to more than the total cost, that charges i more than j if i's increment exceeds j's. Unfortunately, all such procedures are inconsistent with the minimum conditions noted above.

To be precise, suppose $n \geqslant 3$ cities are to be visited in fixed order $(1, 2, \ldots, n)$ before returning home (city 0). Let C be the intercity travel cost of this tour, and let C_i be the travel cost if city i is omitted but all others are visited in their original order, so that sponsor i's incremental cost is $C - C_i$. Then there is no general scheme that (1) allocates exactly C among the n sponsors, (2) charges no sponsor more than the direct round-trip cost between home and that sponsor's city, and (3) charges i more than j when $C - C_i$ exceeds $C - C_j$. Fig. 1 shows why this is so. Since $C - C_2$ exceeds the other $C - C_i$, (1) and (3) require sponsor 2 to pay more than \$500. But this greatly exceeds the \$220 round-trip cost between home and city 2, thus violating (2). Consequently, we must look elsewhere if (1) and (2) are to be honored.

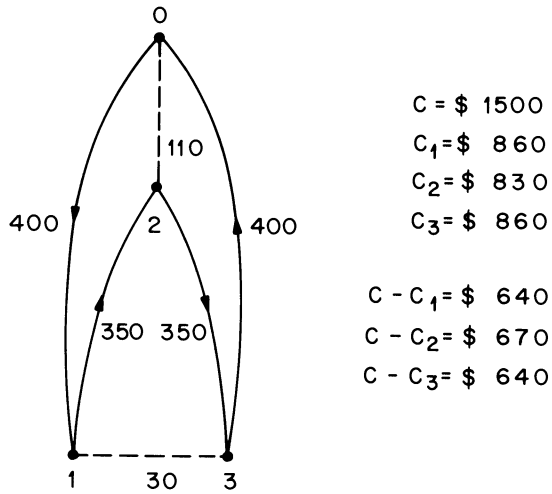


FIG. 1. Incremental costs.

The actual scheme chosen for the tour described in the first paragraph is what we shall call the proportional WTP (willingness to pay) scheme. Proportional WTP charges each sponsor in direct proportion to the cost of going directly from the home city to the sponsor's city. Thus, if cities $1, 2, \ldots, n$ are visited before returning home, if C is the actual air fare, and if c_i is the direct cost of going from home to city i, then $(c_i/\Sigma c_j)C$ is charged to the sponsor at city i. The case described in the opening paragraph is illustrated—without regard to geography—in Fig. 2, where x_i is the cost of going directly from i to $i + 1$.

Although proportional WTP is only one among many allocation schemes for situations of this type, we find it very attractive and will devote most of the paper to it. We will show that simple and appealing conditions for fair allocation in conjunction with either a general separability axiom or an additivity axiom lead to specialized classes of schemes which contain proportional WTP. A further condition that also can be defended as a criterion of fairness then singles out proportional WTP.

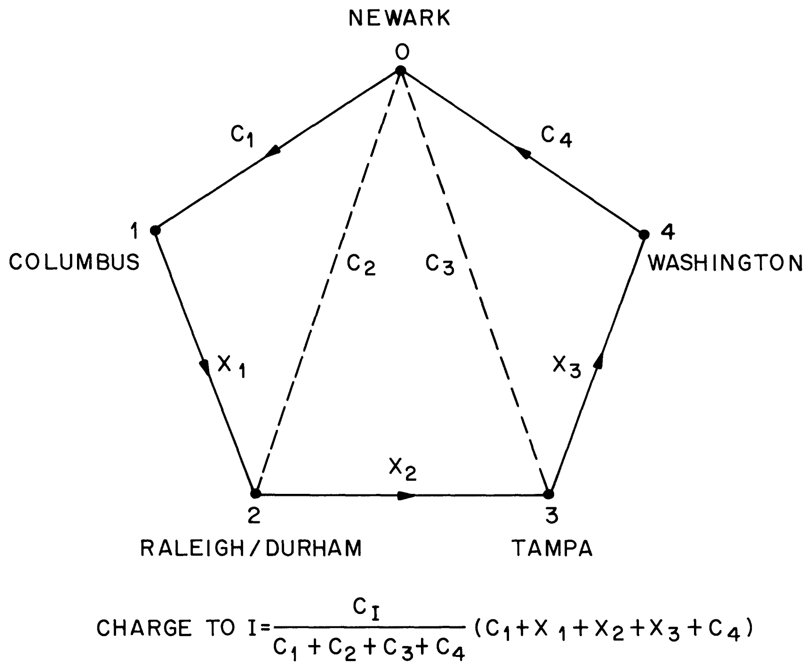


FIG. 2. Proportional WTP scheme.

In connection with Figs. 1 and 2 it should be emphasized that we are not concerned with minimizing total intercity travel costs, as in the familiar traveling salesman problem. Rather, we take the travel route as given, whether it is a minimum-cost route or not, and proceed on the basis of this fixed route. Various factors other than costs can of course affect the choice of route, so to retain flexibility we allow all possibilities that satisfy a version of the triangle inequality for intercity costs.

The fixed-route assumption will be relaxed in the final section, where we mention generalizations of our basic approach. The formulation for this approach and basic fairness criteria are set forth in the next section. We then examine the three-city ($n = 2$) case in detail before extending the analysis to larger n.

Although the context of our allocation problem differs from most settings examined in the literature, many methods proposed for allocating joint project costs to sponsors or departments that benefit from the project are based on proportional and/or incremental schemes. An extensive review with numerous references is provided by Thomas [**13**]. The proportional WTP scheme described above is similar to what is sometimes referred to as SCRB (separable costs-remaining benefits) allocation [**4**], [**9**]. Kalai [**5**] shows how a proportional scheme that resembles our WTP proposal arises from simple axioms when allocation is treated as a bargaining game after the manner of Nash [**8**]. Other game-theoretic approaches to allocation, based in part on Shapley's theory [**10**], are discussed by Shubik [**12**], Sharkey [**11**], Littlechild [**6**] and Young [**14**], among others. Two recent axiomatizations of allocation schemes that are based directly on the factors of allocation without embedding these in a game-theoretic formulation are Billera and Heath [**1**] and Mirman and Tauman [**7**]. Our approach resembles theirs.

2. Formulation. We assume that an individual travels from his home city (city 0) to cities $1, 2, \ldots, n$, in that order, and then returns home. Let x_{ij} denote the cost of traveling directly from city i to city j. It is assumed that all x_{ij} for $i \neq j$ are positive, symmetric ($x_{ij} = x_{ji}$), and obey the triangle inequality

$$x_{ij} + x_{jk} \geqslant x_{ik}.$$

Although violations of this inequality occur, it holds in most cases and will be an integral factor in our analysis. A relaxation is mentioned in the final section.

For convenience, we let $c_i = x_{0i}$, the direct cost from home to city i, with $c = (c_1, \ldots, c_n)$ the vector of these costs. In addition, x_i is defined as $x_{i,i+1}$, with $x = (x_1, \ldots, x_{n-1})$. The individual's intercity travel costs are

$$C = c_1 + \sum_{i=1}^{n-1} x_i + c_n.$$

Since $x_i \leqslant c_i + c_{i+1}$ by the triangle inequality, C never exceeds the sum of the round-trip costs between home and the other cities, $\Sigma 2c_i$. Indeed, $C = 2\Sigma c_i$ only when $x_i = c_i + c_{i+1}$ for all i; a plausible picture for this has the individual alternately flying back and forth over his home city on each intermediate leg of his trip.

There is a sponsor at each city visited who is to pay part of C. Sponsors at different cities could be identical. As in the introductory example, the individual himself could be the sponsor at one or more cities. We identify the sponsor at city i with index i, it being understood that an entity who is the sponsor at several cities pays the sum of the costs charged to those cities.

It seems natural within this formulation to propose that the intercity travel costs charged to i depend only on the matrix $[x_{ij}]$ of intercity costs. We shall in fact assume more than this, namely that the charge to i depends only on c and x. Thus, it is presumed that direct travel costs between non-home cities that are not visited in succession are irrelevant. Although we feel that this is reasonable in view of the fixed-route context, it does rule out certain incremental allocation schemes when $n \geqslant 3$.

Three Fairness Criteria. Henceforth, let $f_i(c, x)$ be the intercity travel costs charged to i. Three conditions on the f_i that we view as basic fairness criteria and that are similar to conditions discussed elsewhere [12], [13] are, for all applicable (c, x),

A1. $f_1(c, x) + \cdots + f_n(c, x) = C$;

A2. $f_i(c, x) \geqslant 0$ *for all* i;

A3. $f_i(c, x) \leqslant 2c_i$ *for all* i.

Suppose first that the individual himself is not a sponsor. Then the full-allocation axiom A1 says that he neither makes a profit ($\Sigma f_i > C$) nor subsidizes his sponsors ($\Sigma f_i < C$) for his travel costs, which is fair both to the sponsors and himself. Condition A2 is a fairness condition among sponsors. Its failure in the context of A1 would mean that some sponsors get "kickbacks" at the expense of other sponsors.

Suppose next that the individual is a sponsor. Then A1 and A2 do not prevent him from getting a free ride to his own sponsor city, but they do imply that he does not "profit" from his travel costs. Even though common practice appears to sanction free rides, they will be ruled out by later conditions. An exception to this appears in the last paragraph of the paper.

Axiom A3 is a willingness-to-pay condition. On an ethical level, it reflects the judgment that it would be unfair to i to charge him more than the round-trip cost between home and city i. Moreover, sponsors may simply be unwilling to pay more than the round-trip costs to their cities.

Since $C \leqslant \Sigma 2c_i$ by the triangle inequality, A1, A2 and A3 are simultaneously feasible. When $x_i = c_i + c_{i+1}$ for $i = 1, \ldots, n-1$, A1 and A3 force $f_i(c, x)$ to equal $2c_i$ for $i = 1, \ldots, n$, so that these two conditions fully determine the allocation at the upper bound on x.

Other Fairness Criteria. We shall impose additional conditions on the f_i as we examine $n = 2$ and then larger n in the next two sections. These conditions reflect notions of neutrality, monotonicity, homogeneity and other potential properties of allocation schemes that, for the most part, can be defended as reasonable fairness criteria.

The $n = 2$ case will be discussed first since it is easily visualized and embodies many features of the general case.

3. Two Cities Visited. Our tale of two cities visited begins with two trips, a month or so apart. The first goes from Atlanta (home) to Boston to Chicago and back home. The second reverses the tour, but the costs remain the same: see Fig. 3. Because of the reversal, Chicago becomes city 1 and Boston becomes city 2 on the second trip. Fair treatment suggests that the named cities ought to be charged the same on each trip. In particular, the charge to Boston for the first trip, $f_1(c_1, c_2, x_1)$, should be the same as the charge to Boston for the second trip, $f_2(c_2, c_1, x_1)$, where $c_1 = 150$ and $c_2 = 100$. This anonymity/neutrality condition [**2**] will be identified as

$$\text{A4.} \quad f_2(c_2, c_1, x_1) = f_1(c_1, c_2, x_1).$$

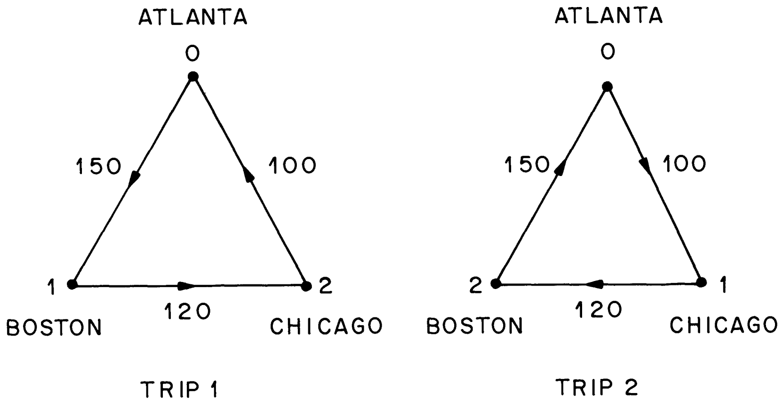


FIG. 3. Trips with opposite tours.

For convenience in the rest of this section we shall sometimes write $f_1(c_1, c_2, x_1)$ as $f(c_1, c_2, x)$. That is, $f(c_1, c_2, x)$ is the charge to sponsor 1 when it costs x to go from 1 to 2. In working towards the proportional WTP scheme, where

$$f(c_1, c_2, x) = \frac{c_1}{c_1 + c_2}(c_1 + x + c_2),$$

we pursue two approaches that delineate classes of schemes which contain proportional WTP. The final step to proportional WTP is taken later in the section.

The First Approach. Our first approach uses only one new condition, a separability axiom for f. As we shall see, this approach is mathematically efficient but lacks some of the intuitive appeal provided by the second approach. The new condition is

A5. *There are real valued functions g on (c, x) pairs, and h and k on c vectors, such that*

$$g(c_1, c_2, x) = g(c_2, c_1, x)$$

and

$$f(c_1, c_2, x) = g(c_1, c_2, x)h(c_1, c_2) + k(c_1, c_2).$$

This says that the charge to 1 consists of a term that is independent of x plus a term that accounts for x in a multiplicative decomposition that is partly symmetric in c_1 and c_2. Although the decomposition of f in A5 retains a great deal of flexibility, our first theorem shows that previous conditions severely limit its possibilities.

Before stating Theorem 1, we remark that if A5 is modified by taking $k \equiv 0$, so that $f(c_1, c_2, x) = g(c_1, c_2, x)h(c_1, c_2)$, then proportional WTP follows directly from the proof of the theorem. In this case $F(c_1, c_2)$ in $(*)$ equals $c_1/(c_1 + c_2)$.

THEOREM 1. *Suppose* A1, A3, A4 *and* A5 *hold. Then there is a nonnegative real valued function F on c vectors such that*

$$F(c_1, c_2) + F(c_2, c_1) = 1$$

and

$$(*) \qquad f(c_1, c_2, x) = 2c_1 - (c_1 + c_2 - x)F(c_1, c_2)$$

for all applicable (c, x) *pairs.*

Expression $(*)$ says that the charge to 1 equals 1's willingness to pay, minus $F(c_1, c_2)$ times the difference between the home-leg costs and the cost of the intermediate leg. Alternatively, for fixed c, 1 pays an amount $2c_1 - (c_1 + c_2)F(c_1, c_2)$, independent of x, plus an amount $xF(c_1, c_2)$ that is linear and nondecreasing in x. Our second approach assumes the monotonicity aspect for x and uses this along with an additivity condition to derive linearity in x.

Given $(*)$, the neutrality condition A4 says that the charge to sponsor 2 is $2c_2 - (c_1 + c_2 - x)F(c_2, c_1)$. Since $F(c_1, c_2) + F(c_2, c_1) = 1$, the sum of the charges is $C = c_1 + c_2 + x$, which agrees with A1. Since $c_1 + c_2 - x \geqslant 0$ by the triangle inequality, and $c_1 + c_2 - x > 0$ for some x, A3 forces F to be nonnegative. Condition A2, which is not used in Theorem 1, is easily seen to be implied by the other conditions. That is, nonnegative charges for the sponsors follow from A1, A3, A4, and A5.

Proof of Theorem 1. By A5

$$f(c_1, c_2, x) = g(c_1, c_2, x)h(c_1, c_2) + k(c_1, c_2),$$
$$f(c_2, c_1, x) = g(c_1, c_2, x)h(c_2, c_1) + k(c_2, c_1).$$

By A1 and A4, $C = f_1(c_1, c_2, x_1) + f_2(c_1, c_2, x_1) = f(c_1, c_2, x) + f(c_2, c_1, x)$. Addition therefore yields

$$c_1 + c_2 + x = g(c, x)[h(c_1, c_2) + h(c_2, c_1)] + k(c_1, c_2) + k(c_2, c_1).$$

If $h(c_1, c_2) + h(c_2, c_1) = 0$, then $c_1 + c_2 + x$ would not depend on x, which is impossible since all costs are assumed positive and x can vary from $|c_1 - c_2|$ to $c_1 + c_2$. We therefore solve the preceding equation for g, and substitute this solution into the first equation of the proof to obtain

$$f(c_1, c_2, x) = (c_1 + c_2 + x)F(c_1, c_2) + G(c_1, c_2),$$

where

$$F(c_1, c_2) = \frac{h(c_1, c_2)}{h(c_1, c_2) + h(c_2, c_1)},$$
$$G(c_1, c_2) = k(c_1, c_2) - [k(c_1, c_2) + k(c_2, c_1)]F(c_1, c_2),$$

with $F(c_1, c_2) + F(c_2, c_1) = 1$. As remarked earlier, A1 and A3 imply $f(c_1, c_2, c_1 + c_2) = 2c_1$, and therefore $2c_1 = 2(c_1 + c_2)F(c_1, c_2) + G(c_1, c_2)$. We solve this for G and substitute into the preceding equation for f to obtain $(*)$. As remarked just before this proof, A3 forces F to be nonnegative. ■

The Second Approach. Our second approach uses the following monotonicity and additivity axioms:

A6. *If* $x \geqslant x'$, *then* $f(c_1, c_2, x) \geqslant f(c_1, c_2, x')$;

A7. $f(c_1, c_2, x) + f(c_1, c_2, x') = f(2c_1, 2c_2, x + x')$.

It would hardly seem fair to 2 to decrease 1's allocation when x increases but the home-to-i costs remain fixed, especially if 2 would have to bear this decrease *plus* the increase in x itself, as would be required under A1. Given A4, the monotonicity axiom A6 says that neither sponsor pays more if the intermediate cost decreases.

Additivity, A7, considers three trips. The first two are the same except for their intermediate costs, x and x'. The third trip doubles the home-to-i costs of the first two and adds the intermediate costs of the first two to obtain the intermediate cost of the third. Indeed, $(2c_1, 2c_2, x$

$+ x') = (c_1, c_2, x) + (c_1, c_2, x')$. Under these hypotheses, it seems reasonable to us that sponsor 1 (likewise 2 under A4) should pay the same for the third trip as he would pay for the other two together.

When $x = x'$ in A7, we obtain the homogeneity property $f(2c_1, 2c_2, 2x) = 2f(c_1, c_2, x)$. This is a special case of general degree-1 homogeneity $[f(\lambda c_1, \lambda c_2, \lambda x) = \lambda f(c_1, c_2, x)]$ for f. Given A1, A3, and A4, which are used in Theorem 1 and in our next theorem, it is easily seen that the conjunction of A5 and $f(2c_1, 2c_2, 2x) = 2f(c_1, c_2, x)$ is equivalent to the conjunction of A6 and A7. Moreover, as shown in the ensuing proof of Theorem 2, A6 and A7 alone imply A5. On the other hand, although A6 is implied by the axioms of Theorem 1, A7 is not since there is nothing in our first approach which requires $F(2c_1, 2c_2)$ to equal $F(c_1, c_2)$. Hence the axioms of the next theorem are stronger than those of Theorem 1, but only because they entail the limited homogeneity property.

THEOREM 2. *Suppose* A1, A3, A4, A6, *and* A7 *hold. Then the conclusions of Theorem* 1 *hold and, in addition,* $F(2c_1, 2c_2) = F(c_1, c_2)$.

Proof. Since we assumed initially that all x_{ij} are positive, we begin the proof of Theorem 2 by supposing that $c_1 \neq c_2$, so that the lower bound $|c_1 - c_2|$ allowed for x by the triangle inequality is feasible. With c_1 and c_2 fixed and $c_1 \neq c_2$, let $x^0 = |c_1 - c_2|$ and $x^1 = c_1 + c_2$. In addition, let $f^0 = f(c_1, c_2, x^0)$ and $f^1 = f(c_1, c_2, x^1)$. By A6, $f^1 \geqslant f^0$.

Let $x^2 = (x^0 + x^1)/2$. Then, by three applications of A7,

$$\begin{aligned} f(c_1, c_2, x^2) &= f(c_1/2, c_2/2, x^0/2) + f(c_1/2, c_2/2, x^1/2) \\ &= \frac{1}{2}f^0 + \frac{1}{2}f^1. \end{aligned}$$

Next, let $x^3 = (x^0 + x^2)/2 = \frac{3}{4}x^0 + \frac{1}{4}x^1$ and $x^4 = (x^1 + x^2)/2 = \frac{1}{4}x^0 + \frac{3}{4}x^1$. Then

$$f(c_1, c_2, x^3) = \frac{1}{2}f^0 + \frac{1}{2}f(c_1, c_2, x^2) = \frac{3}{4}f^0 + \frac{1}{4}f^1,$$

$$f(c_1, c_2, x^4) = \frac{1}{2}f(c_1, c_2, x^2) + \frac{1}{2}f^1 = \frac{1}{4}f^0 + \frac{3}{4}f^1.$$

Further bisections lead to

$$f(c_1, c_2, \alpha x^0 + (1 - \alpha)x^1) = \alpha f^0 + (1 - \alpha)f^1$$

for all α in a subset of $[0, 1]$ that is dense in $[0, 1]$, and monotonicity then forces $f(c_1, c_2, \alpha x^0 + (1 - \alpha)x^1) = \alpha f^0 + (1 - \alpha)f^1$ for all $\alpha \in [0, 1]$. In other words,

$$\begin{aligned} f(c_1, c_2, x) &= \frac{x^1 - x}{x^1 - x^0}f^0 + \frac{x - x^0}{x^1 - x^0}f^1 \\ &= \frac{x^1 f^0 - x^0 f^1}{x^1 - x^0} + x\left(\frac{f^1 - f^0}{x^1 - x^0}\right) \end{aligned}$$

for all $x \in [x^0, x^1]$.

If $c_1 = c_2$, a similar linear form holds for all x in $[\delta, x^1]$, for any $0 < \delta < x^1$. Since this is true no matter how small δ is taken, it follows in general that there are functions h and k on c vectors such that

$$f(c_1, c_2, x) = xh(c_1, c_2) + k(c_1, c_2).$$

Thus, A5 holds, and we obtain the conclusions of Theorem 1 from the hypotheses of Theorem 2. The final conclusion of Theorem 2 follows immediately from (*) and $f(2c_1, 2c_2, 2x) = 2f(c_1, c_2, x)$. ■

Proportional WTP. Henceforth in this section we assume that A1, A3, A4, and either A5 or A6 and A7 hold, so that, with $x = x_1$,

$$f_1(c_1, c_2, x) = 2c_1 - (c_1 + c_2 - x)F(c_1, c_2)$$
$$f_2(c_1, c_2, x) = 2c_2 - (c_1 + c_2 - x)F(c_2, c_1),$$

where $F \geqslant 0$ and $F(c_1, c_2) + F(c_2, c_1) = 1$. There are two situations in which this form completely determines the f_i. The first is $x = c_1 + c_2$, whence $f_i = 2c_i$ for each i. The second arises when $c_1 = c_2$, for then $F(c_1, c_2) = 1/2$, with $f_i = c_i + \frac{1}{2}x$. Thus, given $x < c_1 + c_2$, the sponsors split the cost of the intermediate leg 50-50 when their home-to-i costs are equal. How should they divide x when $c_1 \neq c_2$?

One feasible scheme is to split x to equalize willingness to pay minus actual charge, or, equivalently, to minimize $(2c_1 - f_1)^2 + (2c_2 - f_2)^2$. Then $F \equiv 1/2$ with

$$f_1(c, x) = (3c_1 + x - c_2)/2$$
$$f_2(c, x) = (3c_2 + x - c_1)/2.$$

This gives $f_i \geqslant c_i$ but has other consequences that seem questionable. In particular, if $c_1 < c_2$ and $x = c_2 - c_1$, then 1 pays only c_1 while 2 bears the entire cost of the intermediate leg. Thus, if New York (home) to Detroit costs \$100, Detroit to Dallas costs \$200, and Dallas to New York costs \$300, then the Detroit sponsor pays \$100 and the Dallas sponsor pays \$500. Since this seems less equitable to us than an allocation that charges some of the \$200 to Detroit, such as \$150 for Detroit and \$450 for Dallas, we do not find it very attractive.

Given $c_1 \neq c_2$ and the general solution in terms of F, it should be clear that if $f_1(c_1, c_2, x)$ is specified for any one feasible $x < c_1 + c_2$, then $F(c_1, c_2)$ is determined. For example, the particular solution in the preceding paragraph is fully characterized by $f_1(c_1, c_2, c_2 - c_1) = c_1$ whenever $c_1 < c_2$. In similar fashion, the proportional WTP solution is characterized by $f_1(c_1, c_2, c_2 - c_1) = 2c_1c_2/(c_1 + c_2)$ whenever $c_1 < c_2$, for this gives $F(c_1, c_2) = c_1/(c_1 + c_2)$ and

$$f_1(c_1, c_2, x) = c_1 + \left(\frac{c_1}{c_1 + c_2}\right)x$$
$$= \left(\frac{c_1}{c_1 + c_2}\right)C.$$

Moreover, if $c_1 > c_2$, then $f_1(c_2, c_1, c_1 - c_2) = 2c_1c_2/(c_1 + c_2)$ implies that $F(c_2, c_1) = c_2/(c_1 + c_2)$, and again we get $F(c_1, c_2) = c_1/(c_1 + c_2)$ since $F(c_1, c_2) + F(c_2, c_1) = 1$.

If, given $c_1 < c_2$ and $c_2 - c_1$ as the cost of the intermediate leg, it is generally felt that the most equitable division of this cost between the sponsors is in direct proportion to their c_i, then it follows that proportional WTP is the most equitable particular solution to our general class of F solutions. Although we believe that such a division is quite reasonable, we do not presently have a compelling argument that it is the most equitable division and would invite others to consider the matter.

Instead of using a specific value of $f_1(c_1, c_2, x)$ for $x < c_1 + c_2$ as a basis for selecting proportional WTP from the myriad possibilities of the class of F solutions, one can use a derivative argument at the upper extreme of x, where $x = c_1 + c_2$. We have noted that $f_i = 2c_i$ when $x = c_1 + c_2$. Suppose we agree that for x slightly less than $c_1 + c_2$, each sponsor should pay less than $2c_i$, and that their reductions from $2c_i$ should be approximately proportional to what they pay at $x = c_1 + c_2$, with the proportions becoming exact as x approaches $c_1 + c_2$. Then, since the ratio of their reductions at $x = c_1 + c_2 - \varepsilon$ is $F(c_1, c_2)/F(c_2, c_1)$, and the ratio of what they pay at $x = c_1 + c_2$ is c_1/c_2, we get $F(c_1, c_2)/F(c_2, c_1) = c_1/c_2$ and $F(c_1, c_2) = c_1/(c_1 + c_2)$.

Integration offers a third approach to proportional WTP. If the average charges to the sponsors over all $x \in [|c_1 - c_2|, c_1 + c_2]$ are assumed to be proportional to what they pay when $x = c_1 + c_2$,

then again we find that $F(c_1, c_2) = c_1/(c_1 + c_2)$.

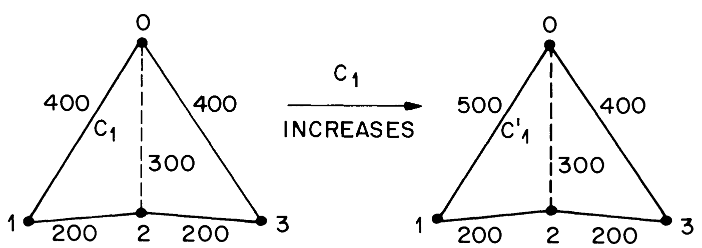
Thus, we see that several distinctly different conditions lead to proportional WTP. However, the ones we have identified are all phrased directly in terms of proportions. Can less obvious and perhaps more convincing conditions be found?

4. Many Cities. Several interesting features of our basic allocation problem appear only when $n \geqslant 3$. One of these is the possibility that C is substantially smaller than Σc_i so that, unlike the case for $n = 2$, we do not think of assigning c_i to sponsor i as his minimum charge. Related to this is the obvious fact that the individual does not travel directly between home and city i for $2 \leqslant i \leqslant n - 1$, so that there might be a temptation to treat 1 and n differently than the intervening sponsors. If one x_i increases but everything else remains fixed, should the f_j for j not adjacent to i change? If an intermediate c_t increases but all else, including C, is unchanged, should the f_t change in specified directions?

As explained earlier, our present discussion is motivated by the proportional WTP scheme in which the allocation to sponsor i is

$$f_i(c, x) = \frac{c_i}{c_1 + c_2 + \cdots + c_{n-1} + c_n}(c_1 + x_1 + \cdots + x_{n-1} + c_n).$$

In this case an increase in any x_i will increase all f_j, an increase in c_i for $2 \leqslant i \leqslant n - 1$ will increase f_i and decrease f_j for all $j \neq i$ (C remaining unchanged), while an increase in c_1 will increase f_1 but can either increase or decrease the other f_i depending on the sign of $\Sigma c_i - C$. Fig. 4 presents a case in which a \$100 increase in c_1 increases f_1 by more than \$100 and decreases each of f_2 and f_3 by a few dollars.



$f_1 = 436$ $\qquad$ $f'_1 = 542$

$f_2 = 327$ $\qquad$ $f'_2 = 325$

$f_3 = 436$ $\qquad$ $f'_3 = 433$

FIG. 4. A change in c_1; proportional WTP.

Two Approaches. Our analysis of the n-city case will be similar to the $n = 2$ analysis of the preceding section. In place of A4 and A5 for the first approach for $n = 2$, we now use a single separability condition which directly involves all sponsors. Two versions of the separability condition will be considered. We state the stronger version first.

A5^0. *There are real valued functions g on (c, x) pairs, and h_i on c vectors for i = 1,..., n, such that*

$$f_i(c, x) = g(c, x)h_i(c) \quad (i = 1, \ldots, n).$$

THEOREM 3. *Axioms* A1, A3 *and* A5^0 *imply proportional WTP.*

Thus, A5^0 is a very strong condition. A weaker version, which retains the flexibility of A5, is

A5*. *There are real valued functions g on (c, x) pairs, and h_i and k_i on c vectors for i = 1,..., n, such that*

$$f_i(c, x) = g(c, x)h_i(c) + k_i(c) \quad (i = 1, \ldots, n).$$

A sense of neutrality is imparted by g, which operates on (c, x) in the same way for all i, and is the only part of the decomposition that involves x.

THEOREM 4. *Suppose* A1, A3 *and* A5* *hold. Then there are nonnegative real valued functions* F_i *on c vectors for* $i = 1, \ldots, n$ *such that*

$$\sum_{i=1}^{n} F_i(c) = 1$$

and

$$(**) \qquad f_i(c, x) = 2c_i - \left(\sum_{j=1}^{n} 2c_j - C\right) F_i(c) \quad (i = 1, \ldots, n)$$

for all applicable (c, x) *pairs.*

Proofs. The proof of Theorem 4 is similar to the proof of Theorem 1. Add the f_i using A5*, solve for $g = [C - \Sigma k_i]/\Sigma h_i$ with the aid of A1, then use the extreme solution at $x_i = c_i + c_{i+1}$ to obtain $(**)$, where $F_i(c) = h_i(c)/\Sigma h_j(c)$. Clearly, $\Sigma F_i = 1$, and $F_i \geqslant 0$ by $(**)$ and A3.

When $k_i \equiv 0$ (Theorem 3), $g = C/\Sigma h_j$, so $f_i(c, x) = (h_i/\Sigma h_j)C$. With each x_i set at $c_i + c_{i+1}$, $f_i = 2c_i = (h_i/\Sigma h_j)(\Sigma 2c_j)$, so $h_i/\Sigma h_j = [c_i/(c_1 + c_2 + \cdots + c_n)]$. ■

As written in Theorem 4, $(**)$ says that the charge to i equals his willingness to pay, minus a fraction of the total difference between all sponsors' willingness to pay and the actual cost C of intercity travel. We noted earlier that $(**)$ yields nonnegative f_i when $n = 2$. However, axiom A2 is not automatically satisfied by the solution in Theorem 4 when $n \geqslant 3$.

Although the axioms used for Theorem 4 do not imply the proportional WTP scheme, they imply the phenomenon illustrated for sponsor 1 in Fig. 4. Assume the conclusions of Theorem 4 hold and, in addition, suppose that an increase of δ in c_1 (or c_n)—all else unchanged—does not increase f_1 (or f_n) by more than δ. That is, we propose, $\partial f_1/\partial c_1 \leqslant 1$ and $\partial f_n/\partial c_n \leqslant 1$. Since

$$\partial f_1/\partial c_1 = 2 - \left[\Sigma 2c_j - C\right] \partial F_1(c)/\partial c_1 - F_1(c),$$

and since C can be made to equal $\Sigma 2c_j$, $\partial f_1/\partial c_1 \leqslant 1$ implies that $F_1(c) \geqslant 1$. Similarly, $\partial f_n/\partial c_n \leqslant 1$ implies $F_n(c) \geqslant 1$. But $F_1 + F_n \geqslant 2$ clearly contradicts the conclusions of the theorem.

Our second approach for the n-city case uses axioms that are similar to A6 and A7. These axioms do not yield as specific a form as $(**)$ and hence do not imply A5* when $n \geqslant 3$, but neither are they implied by A5* since the homogeneity implication of A7*—that $f_i(2c, 2x) = 2f_i(c, x)$—does not follow from A5*. In A6*, $x \geqslant x'$ means that $x_i \geqslant x_i'$ for all $i \leqslant n - 1$:

A6*. *If* $x \geqslant x'$, *then* $f_i(c, x) \geqslant f_i(c, x')$ *for* $i = 1, \ldots, n$;

A7*. $f_i(c, x) + f_i(c, x') = f_i(2c, x + x')$ *for* $i = 1, \ldots, n$.

THEOREM 5. *Suppose* A1, A3, A6* *and* A7* *hold. Then there are nonnegative real valued functions* F_{ij} *on c vectors for* $i = 1, \ldots, n$ *and* $j = 1, \ldots, n - 1$, *such that*

$$\sum_{i=1}^{n} F_{ij}(c) = 1 \quad (j = 1, \ldots, n - 1)$$

and

$$f_i(c, x) = 2c_i - \sum_{j=1}^{n-1} (c_j + c_{j+1} - x_j) F_{ij}(c) \quad (i = 1, \ldots, n)$$

for all applicable (c, x) *pairs.*

For $n \geqslant 3$, the conclusions of Theorem 5 imply those of Theorem 4 only if $F_{i1} = F_{i2} = \cdots =$

$F_{i,n-1}$ for all i. The different forms for f_i in Theorems 4 and 5 will be dealt with further in the next subsection. A proof of Theorem 5 will be presented elsewhere [3].

Proportional WTP. We conclude our examination of the basic case by noting how the proportional WTP scheme follows from the forms for f_i in Theorems 4 and 5 through a derivative condition at the upper x extreme.

Recall that x_i is the cost of going from city i to city $i+1$, for $1 \leqslant i \leqslant n-1$. Suppose that, beginning at $x_j = c_j + c_{j+1}$ for each j, the marginal reductions in the charges to i and $i+1$ that accrue from a reduction in x_i are proportional to what i and $i+1$ paid initially:

$$\left.\frac{\partial f_i(c,x)/\partial x_i}{\partial f_{i+1}(c,x)/\partial x_i}\right|_{\substack{\text{all } x_j = \\ c_j + c_{j+1}}} = \left.\frac{f_i(c,x)}{f_{i+1}(c,x)}\right|_{\substack{\text{all } x_j = \\ c_j + c_{j+1}}} \qquad (i = 1,\ldots,n-1).$$

Given A1, A3, and A5*, this implies proportional WTP. For, given (* *), our derivative condition says that

$$\frac{F_i(c)}{F_{i+1}(c)} = \frac{c_i}{c_{i+1}} \quad \text{for } i = 1,\ldots,n-1,$$

and this along with $\Sigma F_i = 1$ implies that $F_i(c) = c_i/\Sigma c_j$ for $i = 1,\ldots,n$.

The preceding derivative condition does not imply proportional WTP in the context of Theorem 5, since there it implies $F_{ii}(c)/F_{i+1,i}(c) = c_i/c_{i+1}$, which is not sufficient to conclude that $F_{ij}(c) = c_i/\Sigma c_k$ for all i and j. However, if we assume that, beginning at $x_j = c_j + c_{j+1}$ for each j, the marginal savings to i and $i+1$ that accrue from a reduction in *any* x_j are proportional to what i and $i+1$ paid initially, then proportional WTP does follow from the form for f_i in Theorem 5. This stronger derivative condition gives $F_{ij}(c)/F_{i+1,j}(c) = c_i/c_{i+1}$ for all $i, j \in \{1,\ldots,n-1\}$, and the use of $\Sigma_i F_{ij} = 1$ then yields $F_{ij}(c) = c_i/\Sigma c_k$ for all i and j.

Proportional WTP also follows from the axioms of Theorem 5 (A1, A3, A6*, A7*) and the weaker derivative condition provided that we augment these axioms with an assumption that implies $F_{ij}(c) = F_{ik}(c)$ for all relevant i, j, and k. The most obvious assumption for this purpose is $f_i(c,x) = f_i(c,x')$ whenever x and x' are identical in all except two places, say j and k, where $x_j + x_k = x'_j + x'_k$. The form for f_i in Theorem 5 then gives $F_{ij}(c) = F_{ik}(c)$, so that that form reduces to (* *).

5. Summary and Generalizations. We have seen that conditions for equitable allocation—including full-allocation, willingness-to-pay, monotonicity and additivity—imply that the charge to sponsor i is

$$f_i(c,x) = 2c_i - \sum_{j=1}^{n-1}(c_j + c_{j+1} - x_j)F_{ij}(c)$$

with $F_{ij} \geqslant 0$ and $\Sigma_i F_{ij}(c) = 1$ for each j. Here $c = (c_1,\ldots,c_n)$, $x = (x_1,\ldots,x_{n-1})$, c_i is the direct cost from home to city i, and x_i is the cost from city i to city $i+1$. A weak separability axiom in place of monotonicity and additivity yields the alternative form

$$f_i(c,x) = 2c_i - (\Sigma_j 2c_j - C)F_i(c),$$

where $C = c_1 + x_1 + x_2 + \cdots + x_{n-1} + c_n$, the total travel cost. An additional boundary-value condition implies

$$f_i(c,x) = \left(\frac{c_i}{\Sigma c_j}\right)C,$$

the proportional WTP scheme.

We conclude with brief remarks on three generalizations, the first two of which are analyzed in

greater detail in [3]. We take $n \geqslant 3$, as in section 4, and let $x = [x_{ij}]$, as in section 2. It is presumed that allocations depend only on c and the x_{ij} involved in the tour under consideration.

(i). Suppose first that the triangle inequality for costs is replaced by the less-demanding inequality

$$c_1 + (x_{12} + x_{23} + \cdots + x_{n-1,n}) + c_n \leqslant \Sigma 2c_i.$$

Then, because of the greater variety in x that is allowed by this relaxation, the axioms of Theorem 5 imply the stronger conclusion of Theorem 4. In addition, the final step to proportional WTP follows from the appealing condition that $f_i(c, x) = c_i$ for all i when $C = \Sigma_i c_i$.

(ii). Suppose next that the triangle inequality applies and all $n!$ tours are considered. A *tour* is a permutation σ on $\{1, 2, \ldots, n\}$: $\sigma(i)$ is the ith city visited in tour σ. The cost of σ is

$$C(\sigma, x) = c_{\sigma(1)} + \sum_{i=1}^{n-1} x_{\sigma(i)\sigma(i+1)} + c_{\sigma(n)},$$

and the charge to i for tour σ is $f_i^\sigma(c, x)$. Here we adopt an axiom which says that the set of minimum-cost tours is precisely the Pareto optimal set, thus providing an incentive for sponsors to adopt a least-cost tour. When this axiom is applied to the σ-dependent conclusions of Theorem 5, it follows that there are nonnegative $F_i(c)$ which sum to 1 such that

$$f_i^\sigma(c, x) = 2c_i - F_i(c)\left[\Sigma_j 2c_j - C(\sigma, x)\right]$$

for all applicable σ, i, c and x. Note in particular that F_i does not depend on σ. Hence if proportional WTP holds for at least one tour, then it holds for all tours.

(iii). Finally, consider the situation of the preceding paragraph with $2c_i$ replaced by $w_i \geqslant 0$ as sponsor i's indicated willingness to pay. We assume that Σw_i exceeds a specified minimum value that depends on the c_i, and that the triangle inequality holds. Let

$$W = \Sigma w_i \quad \text{and} \quad T(x) = \{\sigma\colon C(\sigma, x) \leqslant W\}.$$

Then $T(x)$ is the set of cost-recoverable tours when w and x apply. We work only with $\sigma \varepsilon T(x)$.

This case breaks down into two subcases, according to whether $W \leqslant \Sigma 2c_i$ or $W > \Sigma 2c_i$. The first of these is amenable to modifications of the conditions used in generalization (ii), which lead to

$$f_i^\sigma(c, x) = \frac{w_i}{W} C(\sigma, x)$$

as the appropriate proportional WTP form. Similar modifications for the "generous-sponsors" subcase, in which $W > \Sigma 2c_i$, yield the less-specific form

$$f_i^\sigma(c, x) = F_i(c) C(\sigma, x), \quad F_i(c) \leqslant w_i / \Sigma 2c_j.$$

Although proportional WTP is feasible for this subcase, it is not uniquely identified by the types of axioms that characterize proportional WTP for the other generalizations.

References

1. L. J. Billera and D. C. Heath, Allocation of shared costs: A set of axioms yielding a unique procedure, Mathematics of Operations Research, 7 (1982) 32–39.
2. D. E. Campbell and P. C. Fishburn, Anonymity conditions in social choice theory, Theory and Decision, 12 (1980) 21–39.
3. P. C. Fishburn and H. O. Pollak, Proportional allocation schemes, mimeographed, Bell Laboratories, 1982.
4. L. D. James and R. R. Lee, Economics of Water Resource Planning, McGraw-Hill, New York, 1971.
5. E. Kalai, Proportional solutions to bargaining situations: interpersonal utility comparisons, Econometrica, 45 (1977) 1623–1630.
6. S. C. Littlechild, Common costs, fixed charges, clubs and games, Review of Economic Studies, 42 (1975) 117–124.
7. L. J. Mirman and Y. Tauman, Demand compatible equitable cost sharing prices, Mathematics of Operations

Research, 7 (1982) 40–56.

8. J. F. Nash, The bargaining problem, Econometrica, 18 (1950) 155–162.

9. J. S. Ransmeier, The Tennessee Valley Authority: A Case Study in the Economics of Multiple Purpose Stream Planning, Vanderbilt Univ. Press, Nashville, Tenn., 1942.

10. L. S. Shapley, A value for n-person games, in H. W. Kuhn and A. W. Tucker (Editors), Contributions to the Theory of Games, vol. II, Princeton Univ. Press, Princeton, N.J., 1953, pp. 307–317.

11. W. W. Sharkey, Suggestions for a game-theoretic approach to public utility pricing, Bell Laboratories Economic Discussion Paper 61, 1974.

12. M. Shubik, Incentives, decentralized control, the assignment of joint costs and internal pricing, Management Science, 8 (1962) 325–343.

13. A. L. Thomas, A Behavioral Analysis of Joint-Cost Allocation and Transfer Pricing, Stipes Publishing Co., 1980.

14. H. P. Young, Cost allocation and demand revelation in public enterprises, International Institute of Applied Systems Analysis Working Paper WP-80-130, 1980.

DO SYMMETRIC PROBLEMS HAVE SYMMETRIC SOLUTIONS?

WILLIAM C. WATERHOUSE

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

There are many problems where the function under study is symmetric in several variables and its maximum or minimum occurs when the variables are equal. Here are some examples:

(1) Of all rectangles with given perimeter, the square has the largest area.

(2) The triangle of minimum area circumscribed about a given circle is equilateral.

(3) Take four positive numbers whose product is 16. Then their sum is at least 8. (That is, the sum is least when all four numbers are equal.)

(4) A box is to be made from sheet metal, with rectangular bottom and sides but no top. Then the design using least material has a square bottom and height half the width. (To reduce this to a symmetric situation, put another copy of the box upside down on top of it.)

(5) Determine the minimum value of

$$(r-1)^2+\left(\frac{s}{r}-1\right)^2+\left(\frac{t}{s}-1\right)^2+\left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \leqslant r \leqslant s \leqslant t \leqslant 4$. (This is a problem from the 1981 Putnam Competition; the minimum comes when $(r, s, t) = (\sqrt{2}, 2, 2\sqrt{2})$. The problem becomes symmetric if we introduce new variables r, s/r, t/s, and $4/t$.)

(6) For a given mean $\bar{x} = (\Sigma_1^n x_i)/n$, the value $\Sigma_1^n x_i^2$ is least when all x_i are equal.

(7) For any positive $x_1, \ldots, x_n$ we have $(\Sigma_1^n x_i)/n \geqslant (x_1 \cdots x_n)^{1/n}$ (the inequality of arithmetic and geometric means, a generalization of (3) above).

It seems clear that some general principle must be lurking here, and I propose to call it the Purkiss Principle. This name honors one of the authors who (independently) noticed this principle in the middle of the last century. Their discussions were somewhat inconclusive and seem to have been completely forgotten; I have not come across any recent references to the idea. Probably this is because the principle is not true without qualifications. But still, there is something to it.

William C. Waterhouse studied at Harvard, receiving his Ph.D. in 1968 under the guidance of John Tate. He taught at Cornell until 1975, when he moved to Penn State. His research has been mainly in algebra, algebraic geometry, and algebraic number theory. He has also published an *Introduction to Affine Group Schemes* and several articles on the history of mathematics.

This work was supported in part by the National Science Foundation, Grant MCS 8102697.

Research, 7 (1982) 40–56.
8. J. F. Nash, The bargaining problem, Econometrica, 18 (1950) 155–162.
9. J. S. Ransmeier, The Tennessee Valley Authority: A Case Study in the Economics of Multiple Purpose Stream Planning, Vanderbilt Univ. Press, Nashville, Tenn., 1942.
10. L. S. Shapley, A value for n-person games, in H. W. Kuhn and A. W. Tucker (Editors), Contributions to the Theory of Games, vol. II, Princeton Univ. Press, Princeton, N.J., 1953, pp. 307–317.
11. W. W. Sharkey, Suggestions for a game-theoretic approach to public utility pricing, Bell Laboratories Economic Discussion Paper 61, 1974.
12. M. Shubik, Incentives, decentralized control, the assignment of joint costs and internal pricing, Management Science, 8 (1962) 325–343.
13. A. L. Thomas, A Behavioral Analysis of Joint-Cost Allocation and Transfer Pricing, Stipes Publishing Co., 1980.
14. H. P. Young, Cost allocation and demand revelation in public enterprises, International Institute of Applied Systems Analysis Working Paper WP-80-130, 1980.

DO SYMMETRIC PROBLEMS HAVE SYMMETRIC SOLUTIONS?

WILLIAM C. WATERHOUSE
Department of Mathematics, The Pennsylvania State University, University Park, PA 16802

There are many problems where the function under study is symmetric in several variables and its maximum or minimum occurs when the variables are equal. Here are some examples:

(1) Of all rectangles with given perimeter, the square has the largest area.

(2) The triangle of minimum area circumscribed about a given circle is equilateral.

(3) Take four positive numbers whose product is 16. Then their sum is at least 8. (That is, the sum is least when all four numbers are equal.)

(4) A box is to be made from sheet metal, with rectangular bottom and sides but no top. Then the design using least material has a square bottom and height half the width. (To reduce this to a symmetric situation, put another copy of the box upside down on top of it.)

(5) Determine the minimum value of

$$(r-1)^2+\left(\frac{s}{r}-1\right)^2+\left(\frac{t}{s}-1\right)^2+\left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \leqslant r \leqslant s \leqslant t \leqslant 4$. (This is a problem from the 1981 Putnam Competition; the minimum comes when $(r, s, t) = (\sqrt{2}, 2, 2\sqrt{2})$. The problem becomes symmetric if we introduce new variables r, s/r, t/s, and $4/t$.)

(6) For a given mean $\bar{x} = (\Sigma_1^n x_i)/n$, the value $\Sigma_1^n x_i^2$ is least when all x_i are equal.

(7) For any positive $x_1, \ldots, x_n$ we have $(\Sigma_1^n x_i)/n \geqslant (x_1 \cdots x_n)^{1/n}$ (the inequality of arithmetic and geometric means, a generalization of (3) above).

It seems clear that some general principle must be lurking here, and I propose to call it the Purkiss Principle. This name honors one of the authors who (independently) noticed this principle in the middle of the last century. Their discussions were somewhat inconclusive and seem to have been completely forgotten; I have not come across any recent references to the idea. Probably this is because the principle is not true without qualifications. But still, there is something to it.

William C. Waterhouse studied at Harvard, receiving his Ph.D. in 1968 under the guidance of John Tate. He taught at Cornell until 1975, when he moved to Penn State. His research has been mainly in algebra, algebraic geometry, and algebraic number theory. He has also published an *Introduction to Affine Group Schemes* and several articles on the history of mathematics.

This work was supported in part by the National Science Foundation, Grant MCS 8102697.

The first person to notice the general pattern behind the specific results was a Frenchman, O. Terquem [13]; the year was 1840. Terquem, who was then nearly sixty years old, was more a man of scholarship than a creative mathematician, and he is perhaps best remembered for his Voltairean credo, "I count all honest men as true believers, and only scoundrels as heretics." (More information about him and our other authors can be found in the biographical appendix below.) Terquem's paper was just a one-page note, and it was prompted by inequalities (6) and (7) above, which occur (at different places) in Cauchy's *Cours d'Analyse*. He considered the general principle obvious, saying that if the numbers x_i "all remain positive, and a given symmetric function preserves a given value, it is evident that every other symmetric function... will attain an extreme value, maximum or minimum, when all the [numbers] become equal." His main concern then was to distinguish the cases of maximum and minimum, which he did by considering the values that occur when one or more of the numbers go to zero.

Terquem's reaction seems to be a very natural one. That is, as soon as the symmetry of a problem is brought out, people are inclined to say that "by symmetry" the extreme value must occur when the variables are equal. But a bit of thought shows that there is no simple symmetry argument to this effect. Indeed, there cannot be, because such a symmetry conclusion is sometimes false. This was the discovery of our next author, the Russian V. Bouniakovsky [2], in 1854. Unlike the other two, Bunyakovskii (to give another common transliteration of his name) was a major mathematician, and there is no problem with his results. He looked at a single symmetric polynomial, asking whether its maxima or minima (if such exist) occur at points where the variables are equal. For polynomials of degree at most three he proved this is true, but in higher degrees he gave counterexamples. His basic example is neat and simple,

$$f(x, y) = \left[x^2 + (y-1)^2\right]\left[(x-1)^2 + y^2\right].$$

Obviously this is symmetric, and obviously also it has its minimum value at $(1,0)$ and $(0,1)$ rather than at any point where $x = y$.

Bouniakovsky did not deal directly with the constrained extremum problems that were Terquem's concern (and ours), but it is not too hard to find counterexamples in that context as well. Take for instance the symmetric function

$$f(x, y) = \left(x^2 + y^2 - 5/8\right)^2,$$

and consider its values for positive x and y satisfying $x + y = 1$. At the equality point $x = y = 1/2$ we get $f(1/2, 1/2) = (1/8)^2$. But f also takes on values larger and smaller than this, for instance $f(1/8, 7/8) = (5/32)^2$ and $f(1/4, 3/4) = 0$. Thus it would seem that our principle simply isn't true.

Still, it is hard to believe that all our original examples had symmetric extrema just by chance. It is time, therefore, that we turned to our third author, the Englishman H. J. Purkiss, who published his observation in 1862 in the *Messenger of Mathematics* [9]. This was then a new journal designed for English undergraduates, and Purkiss was one of the founders; at that time he was in fact a student at Trinity College, Cambridge, and was just twenty years old. Unlike Terquem, he realized that his observation required proof. Needless to say, his argument was unsatisfactory, but it has one significant feature: it deals only with small variations away from equality of the variables. Though Purkiss himself slurred over the distinction between relative and absolute maxima, his argument is the first indication of a basic fact: whatever validity the principle has will be *local*. For example, our function $(x^2 + y^2 - 5/8)^2$ on the line $x + y = 1$ does indeed have a local maximum at $x = y = 1/2$, even though it reaches greater values far away from that point. Thus we can now give a more precise formulation:

THE PURKISS PRINCIPLE. *Let $f(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n)$ be two symmetric functions. On the set where g stays equal to $g(r, \ldots, r)$, the function f should have a local maximum or minimum at $(r, \ldots, r)$.*

Of course we do not yet have any idea why, or indeed whether, the principle should be true. All we know (from Bouniakovsky) is that no rudimentary symmetry argument will prove it. But at least we can try testing it by the general method to test for a constrained extremum at a point P. The classical way to remember this "Lagrange multiplier" method is to think of curves $c(t)$ that have $c(0)=P$ and lie in the set where g is constant. Differentiating the condition $g(c(t))=$ constant, we find that

$$0=\Sigma(D_i g)(P)c_i'(0),$$

which says that the curve's tangent vector $c'(0)$ is perpendicular to the gradient vector $(\nabla g)(P)$ formed by the $D_i(g)=\partial g/\partial x_i$. Conversely, so long as $(\nabla g)(P)$ is nontrivial, the implicit function theorem shows that every vector perpendicular to $(\nabla g)(P)$ occurs as the tangent to some such curve $c(t)$. Now if f has a local (constrained) extremum at P, each $f(c(t))$ has an extremum at 0, and its derivative vanishes there. Hence we have

$$0=\Sigma(D_i f)(P)c_i'(0).$$

This says that $(\nabla f)(P)$ is also perpendicular to all the $c'(0)$, which implies that

$$(\nabla f)(P) \text{ is a scalar multiple of } (\nabla g)(P).$$

Points where this condition holds (together with points where $\nabla g=0$) are the critical points for the constrained extremum problem.

All this is familiar in general. Does something special happen when f and g are symmetric? Let us work out an example, say

$$f(x_1,x_2,x_3)=x_1x_2+x_2x_3+x_3x_1,$$
$$g(x_1,x_2,x_3)=(x_1x_2x_3)^2.$$

The gradients of these are

$$\nabla f=(x_2+x_3,x_1+x_3,x_1+x_2),$$
$$\nabla g=2x_1x_2x_3(x_2x_3,x_1x_3,x_1x_2).$$

We are interested in points P of the form (r,r,r), and there we have

$$\nabla f(r,r,r)=(2r,2r,2r)\quad=2r(1,1,1),$$
$$\nabla g(r,r,r)=(2r^5,2r^5,2r^5)=2r^5(1,1,1).$$

Thus ∇f is a multiple of ∇g simply because in each of them all the entries are equal. Once we see that this is what we want to prove, we have no difficulty proving it.

LEMMA 1. *Suppose $f(x_1,\dots,x_n)$ is a symmetric differentiable function. Then at a point $x_1=\cdots=x_n=r$, all the D_if are equal.*

Proof. Let π be a permutation of $1,2,\dots,n$. For any function h we can define

$$h_\pi(x_1,\dots,x_n)=h(x_{\pi(1)},\dots,x_{\pi(n)}),$$

and trivially then

$$(D_{\pi(i)}h_\pi)(x_1,\dots,x_n)=(D_ih)(x_{\pi(1)},\dots,x_{\pi(n)}).$$

In the present case all f_π equal f and all x_i are equal. ■

With hindsight (and a little charity) we might say that Purkiss also managed to show that the points $(r,\dots,r)$ are critical points. But then he undeniably fell into an error. When there is only one free variable, a critical point is "in general" a local maximum or minimum—that is, it will be one or the other except for degenerate cases where the second derivative vanishes. But in several variables this is not true, and nondegenerate critical points can equally well be saddle points.

Purkiss simply ignored this possibility, and thus he left a major gap in his argument. Yet it must be admitted that in our original examples we did not in fact encounter any saddle points. Support for the principle can also be drawn from George Chrystal, who in 1889 included it in Part 2 of his famous textbook on algebra [**4**, II. 61–63]. Recognizing the inadequacy of Purkiss' proof, Chrystal treated only the case where f and g are symmetric polynomials; these he rewrote in terms of the elementary symmetric polynomials, and after some computation he was able to establish the principle (apart from some degenerate situations). With this encouragement, then, let us take up where Purkiss left off and try to show that for some reason we never get saddle points.

For this we need the second-order terms in the Lagrange multiplier method. These ought to be familiar, but in fact most advanced calculus books seem to skip them. Briefly, then, let us again consider one of our curves $c(t)$ lying in the set where $g = g(P)$. We have

$$\begin{aligned} 0 &= \left.\frac{d^2}{dt^2} g(c(t))\right|_{t=0} \\ &= \Sigma_i (D_i g)(P) c_i''(0) + \Sigma_{i,j}(D_i D_j g)(P) c_i'(0) c_j'(0) \end{aligned}$$

and

$$\frac{d^2}{dt^2} f(c(t))|_{t=0} = \Sigma_i (D_i f)(P) c_i''(0) + \Sigma_{i,j}(D_i D_j f)(P) c_i'(0) c_j'(0).$$

We know that at our critical point we have $\nabla f(P) = \lambda(\nabla g)(P)$ for some scalar λ. Multiplying the first equation above by λ and subtracting, we get

$$\frac{d^2}{dt^2} f(c(t))|_{t=0} = \Sigma_{i,j}\left[(D_i D_j f)(P) - \lambda(D_i D_j g)(P)\right] c_i'(0) c_j'(0).$$

For a local maximum or minimum, we want these second derivatives to have the same sign for all $c(t)$. Thus the extra condition we need is that the quadratic form

$$Q(v) = \Sigma\left[(D_i D_j f)(P) - \lambda(D_i D_j g)(P)\right] v_i v_j$$

should be positive definite or negative definite on the space of all v perpendicular to $\nabla g(P)$. Using the implicit function theorem, one can show that this condition is indeed sufficient [**5**, 154].

This tells us what we need, but why should it automatically be true for symmetric f and g? Let us look again at our example,

$$f = x_1 x_2 + x_2 x_3 + x_3 x_1, \quad g = (x_1 x_2 x_3)^2.$$

At a point (r, r, r) we find that the matrices of second partials are

$$\left(D_i D_j f(P)\right) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

$$\left(D_i D_j g(P)\right) = \begin{pmatrix} 2r^4 & 4r^4 & 4r^4 \\ 4r^4 & 2r^4 & 4r^4 \\ 4r^4 & 4r^4 & 2r^4 \end{pmatrix}.$$

What strikes the eye here is that all the diagonal entries are equal and all the off-diagonal entries are equal. Again, once we notice this we can easily check it in general.

LEMMA 2. *Suppose $f(x_1,\dots, x_n)$ is a symmetric twice-differentiable function. Then at a point $(r,\dots, r)$ all $D_i D_i f$ are equal and all $D_i D_j f$ for $i \neq j$ are equal.*

Proof. In the proof of Lemma 1 we saw that $D_{\pi(i)} h = (D_i h)_\pi$ for any h. Applying this rule twice, we get

$$D_{\pi(i)}D_{\pi(j)}f = D_{\pi(i)}D_{\pi(j)}f_\pi = D_{\pi(i)}\left[(D_jf)_\pi\right] = \left[D_iD_jf\right]_\pi.$$

Take first $i = j$, and choose π so $\pi(i) = 1$; this gives us $(D_1D_1f)(r,\dots,r) = (D_iD_if)(r,\dots,r)$. Take next any $i \neq j$, and choose π so $\pi(i) = 1$ and $\pi(j) = 2$; then we get $(D_1D_2f)(r,\dots,r) = (D_iD_jf)(r,\dots,r)$. ■

A straightforward computation now gives us the final lemma we need:

LEMMA 3. *Suppose a quadratic form Q is given by*

$$Q(v_1,\dots,v_n) = \Sigma_i bv_iv_i + \Sigma_{i\neq j}cv_iv_j.$$

Then for all v satisfying $\Sigma v_i = 0$, we have $Q(v) = (c-b)\Sigma v_i^2$.

THEOREM (The PURKISS PRINCIPLE). *Let f and g be symmetric functions with continuous second derivatives in the neighborhood of a point $P = (r,\dots,r)$. On the set where g equals $g(P)$, the function f will have a local maximum or minimum at P except in degenerate cases.*

Proof. Assuming $\nabla g(P) \neq 0$, we know from Lemma 1 that $\nabla f(P)$ has the form $\lambda\nabla g(P)$. The second partial derivatives of f and g satisfy the equalities in Lemma 2, and hence the terms $(D_iD_jf)(P) - \lambda(D_iD_jg)(P)$ also satisfy those equalities. Lemma 1 shows that the vectors v perpendicular to $\nabla g(P)$ are those with $\Sigma v_i = 0$. Lemma 3 then shows that on those v our quadratic form (if not identically zero) is positive or negative definite. The result then follows from the Lagrange multiplier criterion. ■

Two types of degeneracy were excluded from the proof: we assumed that $\nabla g(P)$ was nonzero, and then we assumed that the second-order terms in $f - \lambda g$ did not all vanish in the directions perpendicular to $\nabla g(P)$. Such exclusions are indeed necessary, and the Purkiss principle is not universally true. To see why a condition on ∇g is needed, consider the extreme case where g is constant everywhere; there is then no constraint, and hardly any symmetric f will satisfy the condition. It is a little harder to find an example that fails with the second-order degeneracy, but here is one such case. Let the functions be

$$f = x_1^4x_2x_3 + x_2^4x_3x_1 + x_3^4x_1x_2,$$

$$g = x_1^3x_2^3 + x_2^3x_3^3 + x_3^3x_1^3$$

around the point $P = (1,1,1)$. Consider the curve consisting of points (st, s, s), where t is close to 1 and s is chosen to keep $g = 3$ (that is, $s^6(2t^3+1) = 3$). For such points

$$f(st,s,s) = s^6(t^4+2t) = 3(t^4+2t)/(2t^3+1).$$

The derivative of this comes out to be $6(t^3-1)^2/(2t^3+1)^2$, which is zero at $t = 1$ but positive on both sides nearby. Thus f has values less than 3 for $t = 1-\varepsilon$ and bigger than 3 for $t = 1+\varepsilon$, and the Purkiss principle fails. It is not at all surprising that such exceptions should exist. What is surprising is how widely the principle turns out to be correct.

Thus far we have come in direct pursuit of the Purkiss principle. By good classical methods we have analyzed it; so far as it is true, we have proved it. Here we might stop. But in fact, the scope of the principle is much wider than we have yet imagined. For an alert reader, the first sign of this should have been Lemma 3. That was a straightforward computation, but in the line of argument it seemed to be a miracle, because for no apparent reason it gave us just what we needed. Such a seeming miracle should never be accepted on its face, for nine times out of ten it marks the presence of a general property not yet recognized or understood.

Furthermore, there are examples to suggest a more general result. Think for instance of a 90° rotation around the vertical axis in three-space. This rotation preserves $x_1^2 + x_2^2 + x_3^2$, and in particular it maps the unit sphere to itself, leaving the north and south poles fixed. Let f be any other function preserved by the same quarter turn, and consider its values on the unit sphere. It is

intuitively clear that the north pole and south pole must be critical points for f on the sphere. Depending on the strength of your intuition, perhaps you can also see that the poles will be extrema rather than saddle points (apart from degenerate cases). Certainly it is easy to check this in examples. But this situation is not covered by the original Purkiss principle.

The geometric language in this example can serve to remind us that symmetry was a geometric concept to begin with. The symmetry that comes from interchangeability of the variables is only one of many types that a function might possess. What elementary geometry suggests is that we should allow other linear changes of variable. (We leave out translations because we are interested in fixed points.) As in all symmetry situations, two changes of variable that preserve a function f can be composed to get another one, and we will be dealing with a group of transformations. There is no reason to bother with continuous families of maps preserving f, since usually then we could reduce f to a function of fewer variables. Thus we are led to consider what happens to the Purkiss principle when we replace interchanges of variables by some other finite group of linear transformations.

Unavoidably now we must raise the level of discussion to assume some familiarity with linear algebra. Our basic structure will be a finite group $\mathbf{G}$ of linear transformations on a finite-dimensional real vector space V. The "symmetric" functions will be those f that satisfy $f(Tv) \equiv f(v)$ for all T in $\mathbf{G}$; the usual name for these is *invariant functions*. The points $(r,\dots,r)$ that we considered before are precisely the ones unaffected by interchanges of coordinates; in our general situation we should correspondingly consider the *fixed points* P in V, those sent to themselves by every T in $\mathbf{G}$. We must also make explicit an invariance property that was hidden in the earlier treatment: interchanging coordinates simultaneously in two vectors $(x_1,\dots,x_n)$ and $(y_i,\dots,y_n)$ does not change the dot product $x \cdot y = \Sigma x_i y_i$. The corresponding tool in general is given to us by the following standard result.

LEMMA 4. *There exists a positive definite inner product* $(\ ,\ )$ *on* V *that is* $\mathbf{G}$*-invariant; in other words,* $(Sv, Sw) = (v, w)$ *for every* S *in* $\mathbf{G}$ *and* v, w *in* V.

Proof. Let $v \cdot w$ be the dot product in some coordinate system. Define

$$(v, w) = \Sigma_{T \text{ in } \mathbf{G}} Tv \cdot Tw.$$

Clearly this is symmetric and bilinear and positive definite. For S in $\mathbf{G}$ we have $(Sv, Sw) = \Sigma TSv \cdot TSw$; and since $\mathbf{G}$ is a group, the terms of this sum are just the terms of (v, w) in different order. ■

Consider now a Taylor expansion of a function f around a point P: it can be written in the form

$$f(P + v) = f(P) + (\nabla f(P), v) + Q_P^f(v) + R(v),$$

where Q_P^f is a quadratic form and $R(v)$ is a remainder term satisfying

$$\lim_{v \to 0} R(v)/(v, v) = 0.$$

In such an expression the vector $\nabla f(P)$ and the quadratic form Q_P^f are uniquely determined. Suppose therefore that we take a fixed point P and an invariant function f. For each T in $\mathbf{G}$ we have then

$$\begin{aligned} f(P + v) &= f(T(P + v)) = f(P + Tv) \\ &= f(P) + (\nabla f(P), Tv) + Q_P^f(Tv) + R(Tv). \end{aligned}$$

By uniqueness we see that $Q_P^f(Tv) = Q_P^f(v)$ and $(\nabla f(P), Tv) = (\nabla f(P), v)$. Invariance of the inner product tells us that $(\nabla f(P), Tv) = (T^{-1}\nabla f(P), v)$; and since this equals $(\nabla f(P), v)$ for all v, we have $T^{-1}\nabla f(P) = \nabla f(P)$. Thus we get the following result, which is as close as we can come to Lemmas 1 and 2 in this generality.

LEMMA 5. *Let f be invariant and P fixed. Then $\nabla f(P)$ is fixed, and the quadratic form Q_P^f is invariant (that is, $Q_P^f(Tv) \equiv Q_P^f(v)$ for all T in* **G**).

Nothing so strong as the Purkiss principle will be true unless we make some further assumptions about our group of transformations. To get the analogue of Lemma 1, for instance, we need to know that the fixed vector $\nabla f(P)$ is forced to be a multiple of the fixed vector $\nabla g(P)$. Thus we should assume that the fixed vectors in V form a one-dimensional subspace. After that, we must separate two types of terms, as in Lemma 2; for this we need an analogue of the space where $\Sigma v_i = 0$. The general concept needed here is that of an *invariant subspace*, one sent to itself by all maps in **G**. We then have the following standard result.

LEMMA 6. *Let U be an invariant subspace. Let W be its orthogonal complement with respect to $(\ ,\ )$. Then W is also an invariant subspace.*

Proof. Take any w in W and T in **G**. Let u be any element of U. Then (Tw, u) equals $(w, T^{-1}u)$, and this is zero because $T^{-1}u$ is in U. Thus Tw is orthogonal to U. ■

In particular, our one-dimensional space U_0 of fixed vectors has an orthogonal complement W_0 which is an invariant subspace. Now if we look back at the proof of Lemma 2, we see that it depended on our having a great many permutations, enough to match most of the matrix entries with each other. The general analogue of this is *irreducibility*, which means that there should be no nontrivial invariant subspaces inside W_0. Using this, we can finally establish the general property behind Lemma 3.

LEMMA 7. *Let* **G** *be a finite group of linear transformations acting irreducibly on a finite-dimensional real vector space W. Let $Q(w)$ be a* **G**-*invariant quadratic form on W. Then $Q(w)$ is a scalar multiple of (w, w). In particular, Q is positive or negative definite whenever it is not identically zero.*

Proof. We can express the quadratic form Q in terms of the invariant inner product, getting $Q(w) = (Bw, w)$ for a unique self-adjoint linear map $B: W \to W$. For each T in **G** we have

$$(Bw, w) = Q(w) = Q(Tw) = (BTw, Tw) = (T^{-1}BTw, w).$$

The uniqueness of B tells us then that $T^{-1}BT = B$. In other words, $BT = TB$. But now the real self-adjoint map B has some real eigenvalue b. Let X be $\{w \text{ in } W | Bw = bw\}$, the corresponding eigenspace. For w in X and T in **G** we have $BTw = TBw = Tbw = bTw$, so Tw is in X. Thus X is an invariant subspace of W. Since it is nontrivial, it must be all of W. Thus $B = b \cdot \mathrm{Id}$, and $Q(w) = (Bw, w) = b(w, w)$. ■

We have now established the analogues of our three initial lemmas. As before, we must rule out degenerate cases: that is, we assume that $\nabla g(P)$ is nonzero and that the second-order terms in $f - \lambda g$ around P do not vanish on the space orthogonal to $\nabla g(P)$ under $(\ ,\ )$. The proof of the original theorem then gives us our wider result.

THEOREM (The EXTENDED PURKISS PRINCIPLE). *Let* **G** *be a finite group of linear transformations of a finite-dimensional real vector space V. Assume that the fixed vectors form a one-dimensional subspace, and that* **G** *acts irreducibly on the complementary subspace. Let f and g be twice-differentiable* **G**-*invariant functions on a neighborhood of a fixed vector P. Then on the set where $g(v) = g(P)$, the function f has a local maximum or minimum at P except in degenerate cases.*

Is this now the ultimate generalization of the Purkiss principle? Not at all. Indeed, Purkiss himself mentioned a similar result where the variables were subject to several independent constraints by different functions $g_1 = g_1(P), \ldots, g_r = g_r(P)$. Other examples will easily come to mind. If we are to do more than just accumulate further cases, we must now reach for more general geometric concepts, those from the rudimentary theory of differentiable manifolds. Assuming familiarity with those, we can examine the theorem with a critical eye, distinguishing

the soul of the principle from its mere outward limbs and flourishes.

First of all, we observe that the actual values of the function g are never mentioned. All that matters is the set M of points where g is equal to $g(P)$. Instead of having g symmetric, then, all we need to require is that our group $\mathbf{G}$ should map M to itself. The condition $\nabla g(P) \neq 0$ now is the familiar condition that allows us to introduce smooth coordinates on M around P. Thus we can absorb that condition by saying that our set M should be a differentiable manifold. (There is of course no need for M to be complete; it may be nothing but a smooth piece around P.) The values of f at points off M are also irrelevant, and all we need for the statement is that f should be defined as a smooth $\mathbf{G}$-invariant function on M.

Furthermore, there is no need for the maps in $\mathbf{G}$ to be induced by linear maps on some ambient space. Linearity will enter automatically, because any smooth map of M to M that fixes P will induce a linear map on the tangent space to M at P. This tangent space in our earlier situation was precisely the orthogonal complement to the span of $\nabla g(P)$, and so this is the space where we should assume irreducibility. Thus we are led to formulate a modernized version of our theorem:

THEOREM (The MODERN PURKISS PRINCIPLE). *Let M be a differentiable manifold. Let $\mathbf{G}$ be a finite group of smooth maps from M to M. Let P be a point in M fixed by $\mathbf{G}$, and let f be a differentiable function on M invariant under $\mathbf{G}$. Assume that the action induced by $\mathbf{G}$ on the tangent space at P is (nontrivial and) irreducible. Then P is a critical point of f; and if this critical point is nondegenerate, it is a local maximum or minimum of f.*

Proof. This theorem shows nicely how much can be done with the machinery of manifolds; for though it is a much more general result, it requires no new ideas in its proof. To begin with, f induces a linear function df_P on the tangent space at P. The kernel of df_P will be $\mathbf{G}$-invariant, and irreducibility thus requires df_P to be zero. (We assume that the action is nontrivial in order to rule out a counterexample in dimension one.) Hence P is a critical point for f. Consequently, the second-order terms of f induce a quadratic form on the tangent space, and clearly this form will be $\mathbf{G}$-invariant. By definition one says that the critical point is nondegenerate when this quadratic form is nondegenerate. In that case Lemma 7 again shows that the form is positive definite or negative definite, and P accordingly gives a local minimum or maximum. ■

This theorem has a quite up-to-date sound, and so with it we can end our pursuit of the Purkiss principle. I do not happen to have seen this final version anywhere before, but I make no great claims for it. More important is that we now see how such a result may stand at the end of a search that began with nothing more than some interchangeable variables in calculus problems.

BIOGRAPHICAL APPENDIX

Olry Terquem ([3], [7], [8]) was born in Metz, France, in 1782. His native language was a form of Yiddish, and his early studies were limited to Hebrew and the Talmud. But the upheaval of the French Revolution happened to bring his older brother in contact with the Jewish community at Coblentz, and a tutor engaged there led Olry into more general studies. His weakness in French made him fail in his first application to the École Polytechnique, but he was admitted on his second attempt (1801). In 1804 he began to teach higher mathematics at the lycée in Mayence, moving in 1811 to a professorship in the artillery school there. In 1814 he was called to Paris as professor attached to the Comité de l'Artillerie; in this post he served as librarian for the Dépôt Central d'Artillerie and also as a general scientific consultant. Exempted from mandatory retirement, he continued active until his death at age 80.

Terquem translated technical works related to artillery and also wrote textbooks on algebra, geometry, and mechanics. Of his discoveries in pure mathematics, we might single out his computation of the number of normals from a given point to an algebraic surface of given degree. His mastery of languages helped make him an authority on the history of mathematics. He also wrote many articles on Jewish concerns, advocating for instance prayers in the vernacular

languages rather than Hebrew. Perhaps his most important contribution to mathematics began in 1842, when he was sixty: he was cofounder of the journal *Nouvelles Annales de Mathématiques*, which he continued to edit until his death. Characteristic of the man are these lines from one of his very last letters: "I believe that human intelligence approaches the divine intelligence *asymptotically*. Let us hope!"

Viktor Yakovlevich Bunyakovskii (1804–1889) is a much better known mathematician than the others, and hence less needs to be said about him here [**6**]. After basic studies in Russia, he received his doctorate in mathematics at Paris in 1825. For most of his career he was a professor at the University of St. Petersburg (now Leningrad) and a member of the St. Petersburg Academy of Sciences; from 1864 to his death he was vice-president of the Academy. He wrote about 150 published works in mathematics and mechanics. His most famous discovery, of course, was the inequality

$$\left(\int_a^b fg\right)^2 \leqslant \left(\int_a^b f^2\right)\left(\int_a^b g^2\right),$$

which he published in 1859. Unfortunately it appeared only in a separate pamphlet written in Russian, and the inequality remained unknown in western Europe until it was rediscovered by Schwarz in 1884.

Henry John Purkiss ([10], [11], [14]) was born in London in 1842. In 1859 he was admitted to Trinity College, Cambridge, as a scholarship student (sizar). In 1862 he became one of the founders and editors of *The [Oxford, Cambridge, and Dublin] Messenger of Mathematics*. He received his B.A. degree in 1864 as the top mathematics student in his class (Senior Wrangler, 1st Smith's Prize). From 1864 to 1865 he was Vice-Principal of the College of Naval Architecture, South Kensington, and in 1865 he was named Principal of the Royal School of Naval Architecture. He drowned on September 17, 1865, in the river Cam. Little more is recorded of his sadly short life, but we can list his published papers:

1. Theorem in maxima and minima, Messenger of Math., 1 (1862) 180–183.
2. Dynamical note, Messenger of Math., 2 (1864) 228.
3. The cardioid, Messenger of Math., 2 (1864) 241–249.
4. Note on the comparative value of Simpson's two rules, and on Dr. Woolley's rule, Naval Architects' Trans., 6 (1865) 48–50.
5. The equation of the tangent, Messenger of Math., 3 (1866) 19–22.
6. Notes on pedal coordinates, Messenger of Math., 3 (1866) 83–88.
7. Notions of infinity derived from gnomonic projection, Messenger of Math., 3 (1866) 171–172.
8. On certain formulas of mensuration, Quart. J. Math., 7 (1866) 235–241.

George Chrystal (1851–1911) was a Scotsman ([**1**], [**12**]). After studies in Aberdeen, he was in residence at Peterhouse, Cambridge, from 1872 to 1875. As a student there, he spent much time with Maxwell in the newly opened Cavendish Laboratory. Many of his friends considered this a "waste of time," since Maxwell's work was not covered in the Tripos examination; even so, Chrystal came out second in his class. He was always inclined toward physics: his first major work was an experimental verification of Ohm's Law, and shortly before his death he did significant work on the analysis of seiches (long-lasting waves in lakes). From 1879 until his death he was Professor of Mathematics at the University of Edinburgh, where he played a major role both in reforming the university curriculum and in raising the level of primary and secondary education in Scotland.

Despite all this, Chrystal will always be best known for his classic algebra textbook, which is still in print. Officially he is present in this Appendix because his book gave the first proof of any

significant form of the Purkiss principle, but in fact he is included because the temptation to quote him is irresistible. Here is a bit from an address he gave a year before Part I of his book appeared:

> The whole teaching consists in example grinding. What should be merely the help to attain the end has become the end itself. The result is that algebra, as we teach it, is neither an art nor a science, but an ill-digested farrago of rules whose object is the solution of examination problems.... The end of all education nowadays is to fit the student to be examined; the end of every examination not to be an educational instrument, but to be an examination which a creditable number of men (however badly taught) shall pass. We reap, but we omit to sow.... The cure for all this evil is to give effect to a higher ideal of education in general, and of scientific education in particular.... It takes the hand of God to make a great mind, but contact with a great mind will make a little mind greater.

Here is part of the Preface to Part II of his book, which is devoted mainly to a careful treatment of power series:

> A practice has sprung up of late (encouraged by demands for premature knowledge in certain examinations) of hurrying young students into the manipulation of the machinery of the Differential and Integral Calculus before they have grasped the preliminary notion of a *Limit*... on which all the meaning and all the uses of the Infinitesimal Calculus are based. Besides being to a large extent an educational sham, this course is a sin against the spirit of mathematical progress.

And here, finally, is part of the famous Preface to Part I:

> The first object I have set before me is to develop Algebra as a science, and thereby to increase its usefulness as an educational discipline. I have also endeavoured so to lay the foundations that nothing shall have to be unlearned.... It becomes necessary, if algebra is to be anything more than a mere bundle of unconnected rules, to lay down generally the fundamental laws of the subject, and to proceed deductively.

Amen.

References

1. J. S. Black and C. G. Knott, Professor George Chrystal, M.A., LL.D., Proc. Roy. Soc. Edinburgh, 32 (1913) 477–503.

2. V. Bouniakovsky, Note sur les maxima et les minima d'une fonction symétrique entière de plusieurs variables, Bull. Classe Phys.-Math. Acad. Imper. Sci. St. Petersburg, 12 (1854) 353–361. [This academy is the predecessor of the current Akademija Nauk SSSR.]

3. M. Chasles, Rapport sur les travaux mathématiques de M. O. Terquem, Nouvelles Ann. de Math., (2), 2 (1863) 241–251.

4. G. Chrystal, Algebra: An Elementary Textbook, Parts I and II, A. and C. Black, Edinburgh, 1886 and 1889.

5. C. H. Edwards, Jr., Advanced Calculus of Several Variables, Academic Press, New York, 1973.

6. A. T. Grigorian, Bunyakovskii, Viktor Yakovlevich; in C. C. Gillespie (Editor), Dictionary of Scientific Biography, vol. 15., Scribner, New York, 1978, pp. 66–67.

7. E. Prouhet, Notice sur la vie et les travaux d'Olry Terquem, Bull. Bibliog. Hist. Biog. Math., 8 (1862) 81–90. [This was issued as part of the Nouvelles Ann. de Math., Series 2, volume 1.]

8. E. Prouhet, Terquem (Olry); in J. Michaud et al., Biographie Universelle, Nouvelle Edition 1854 ff., reprint Akademische Druck- und Verlagsanstalt, Graz, 1970, vol. 41, p. 168.

9. H. J. Purkiss, Theorem in maxima and minima, Messenger of Math., 1 (1862) 180–183.

10. Royal Society of London, Catalogue of Scientific Papers (1800–1863), vol. 5, H. M. Stationery Office, London, 1871.

11. Royal Society of London, Catalogue of Scientific Papers (1864–1873), vol. 8, John Murray, London, 1879.

12. R. Schlapp, George Chrystal; in C. C. Gillespie (Editor), Dictionary of Scientific Biography, vol. 3, Scribner, New York, 1971, pp. 264–265.

13. O. Terquem, Démonstration de deux propositions de M. Cauchy, J. Math. Pures Appl. (Liouville), 5 (1840) 37.

14. J. A. Venn (Editor), Alumni Cantabrigienses, Part II, vol. 5, Cambridge Univ. Press, Cambridge, 1953.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P. Transactions of the Moscow Mathematical Society, 1982, Issue 2. AMS, 1982, iv + 268 pp, $91 (P). Translation of Tom 42 (1981).

General, S*(12-14), L. The Prentice-Hall Encyclopedia of Mathematics. Beverly Henderson West, et al. Prentice-Hall, 1982, xv + 683 pp, $35. [ISBN: 0-13-696013-8] Contains eighty articles written for a secondary school or general audience. Explanations are clear and amply illustrated. Extensive historical information is particularly good. Gives numerous references for further reading. RSK

General, T*(13-16: 1), L*. Mathematics, Its Power and Utility. Karl J. Smith. Brooks/Cole, 1983, xviii + 558 pp, $22.95. [ISBN: 0-534-01190-X] For those wishing to overcome "math anxiety." Text, by experienced author, developed around everyday situations with which readers can identify. Stress on problem formulation and solving. Spiced with cartoons, photographs, charts, tables, newspaper and other ads. Lengthy bibliography on math anxiety. Numerous exercises with answers to half. Stunning Karascope-produced image on front cover. JK

Elementary, T(13). Functions and Graphs. Edwin F. Beckenbach, Michael D. Grady, Irving Drooyan. Wadsworth Pub, 1983, x + 578 pp. [ISBN: 0-534-01180-2] The usual topics of a pre-calculus algebra/trigonometry course are all here, but arranged differently, with different emphasis. Might be an approach that avoids the student impression that "I've seen all this before." AWR

Elementary, T(13: 1). Operational Mathematics for Business, Second Edition. R.C. Pierce, Jr., W.J. Tebeaux. Wadsworth Pub, 1983, xx + 642 pp. [ISBN: 0-534-01235-3] Changes from the First Edition (TR, December 1980) inspired by the 1981 Tax Act. Also new are chapter tests, expanded problem sets and expanded chapters on graphing and statistics. JRG

Precalculus, T(13: 1). Precalculus: A Functional Approach to Algebra and Trigonometry. Peter Evanovich, Martin Kerner. Holden-Day, 1982, xiii + 408 pp, $20.50. [ISBN: 0-8162-2715-2] Preliminaries, notation kept to a minimum (e.g., ϵ avoided for set membership). An early chapter introduces functions, while inequalities and absolute values are next to last. Organization appears patched together, with appendices inserted irregularly, and certain explanations seem unclear. Exercises seem overly elementary and repetitive. Some applications. RB

Precalculus, T(13-14: 1). Algebra & Trigonometry with Applications. Jagdish C. Arya, Robin W. Lardner. Prentice-Hall, 1983, xiv + 633 pp, $23.95. [ISBN: 0-13-021675-5] A straightforward ("all signal, no noise") fairly traditional text, the book is somewhat long on algebra (including such topics as Descartes' Rule of Signs and Cramer's Rule) and somewhat short on trigonometry. Attractive format, good selection of routine exercises and applications. Tables, answers, index. JS

Precalculus, T(13: 1). Basic Mathematics for Calculus, Second Edition. Dennis G. Zill, Jacqueline M. Dewar, Warren S. Wright. Wadsworth Pub, 1983, x + 437 pp. [ISBN: 0-534-01197-7] Informal, "no-nonsense" approach to precalculus topics. Includes numerous examples, calculator problems, chapter reviews and tests, and short appendices for basic math review, mathematical induction, complex numbers and axis rotation. Application problems included in most sections. PB

Precalculus, T(13: 1), S. Mathematics for Technical Education, Second Edition. Dale Ewen, Michael A. Topper. Prentice-Hall, 1983, vii + 504 pp, $23.95. [ISBN: 0-13-565168-9] Essentially a precalculus course which includes transcendental functions, and introductions to vectors, complex numbers, and analytic geometry. New topics in this edition include properties of determinants and inverse trigonometric functions. The text consists primarily of examples and exercises. (First Edition, TR, October 1976.) CEC

Precalculus, T(13-14: 1). College Algebra. Max A. Sobel, Norbert Lerner. Prentice-Hall, 1983, xvi + 435 pp, $22.95. [ISBN: 0-13-141796-7] This text especially emphasizes the algebraic skills that will be used in calculus: simplifying difference quotients, recognizing composite functions (for the chain rule), solving inequalities (for finding signs of derivatives), et cetera. Also stresses graphing (geometric transformations induced by algebraic changes). Points out common pitfalls, has review exercises, and sample tests. GHM

Precalculus, T(13: 1). College Algebra with Applications. Jagdish C. Arya, Robin W. Lardner. Prentice-Hall, 1983, xiv + 508 pp, $22.95. [ISBN: 0-13-140699-X] A review of basic algebra followed by chapters on lines, functions, conics, exponentials and logarithms, polynomials, and matrices and other topics from algebra. Some problems and exercises involve applications. FLW

Precalculus, T*(13: 1). Introduction to Technical Mathematics. Loren Radford, Anthony Vavra, Shirley Rychlicki. Prindle, Weber & Schmidt, 1983, viii + 680 pp. [ISBN: 0-87150-339-5] Readable, student-oriented, comprehensive coverage of linear, quadratic, trigonometric, exponential and logarithmic functions for readers with basic skills in arithmetic, geometry and algebra. Material is related to needs of technologists. Unusual features include chapter on measurements and data handling, and appendices on graphical treatment of data, data analysis and dimensional analysis. Calculator use integrated throughout text. Numerous routine exercises with answers to half. JK

Education, P. The Programmable Hand Calculator: A Teacher's Tool for Mathematics Classroom Lectures. Bernard Seckler. Sigma Pr, 1981, 207 pp, (P). Gives programs, with comments and examples, for both the Hewlett Packard 29C and the Texas Instruments 58C, which can be used to solve several college-level mathematics problems. Topics range from factorization of integers to the Runge-Kutta method for solving a differential equation. RSK

Education, T*(16), P*. Teaching Problem-Solving Strategies. Daniel T. Dolan, James Williamson. Addison-Wesley, 1983, xiv + 322 pp, $16 (P). [ISBN: 0-201-10231-5] Carefully constructed introduction to six problem-solving strategies for integration with junior high math programs. Includes introductory lessons, activities and practice problems with worksheets ready for duplicating. Activities correlated with strategies and standard topics, thoughtful page layouts and interesting mix of topics. PB

Education, S(16-17). Ideas from the Arithmetic Teacher, Grades 6-8, Middle School. George Immerzeel, Melvin Thomas. NCTM, 1982, ii + 142 pp, $5.40 (P). [ISBN: 0-87353-200-7] Reprinted directly from the IDEAS section of this journal. Student activity sheet and teacher suggestions back-to-back on perforated pages for ease of use. Good ideas--ideal way to introduce potential middle school teachers to the Arithmetic Teacher. MW

Education, S(16-17), P. Teaching Mathematics. Ed: Michael Cornelius. Nichols Pub, 1982, 248 pp, $15.00 (P); $27.50. [ISBN: 0-89397-137-5; 0-89397-136-7] Ten essays on current British issues in mathematics education. Practical ideas on providing for exceptional or learning disabled children, teaching through recreational mathematics, and providing applications in other areas. Discusses teaching at primary, middle and secondary level. MW

History, P, L. Loo-Keng Hua Selected Papers. Ed: H. Halberstam. Springer-Verlag, 1983, xiv + 889 pp, $42. [ISBN: 0-387-90744-0] A selection of about 50 papers, supplemented by topical surveys covering number theory, algebra and geometry, and function theory. Includes a brief biography, a complete publication list, and a tantalizing list (without any details) of over 100 industrial problems to which Hua has contributed significantly by leading a team of applied mathematicians in the development of new mathematical methodology. LAS

History, T*(14-16: 2), S, L**.** The Nature and Growth of Modern Mathematics. Edna E. Kramer. Princeton U Pr, 1982, xxiv + 758 pp, $9.95 (P). [ISBN: 0-691-02372-7] A welcome re-issue of one of the best books in its genre: a compelling exposition of mathematics--theories, applications, people, societal context--from antiquity to mid-twentieth century. Fully one-third of the book is devoted to twentieth century mathematics in terms that any serious non-mathematician can comprehend. That in itself makes this book virtually unique. (First Edition, TR, April 1971; Extended Review, April 1974.) LAS

Combinatorics, S(17-18), P. Lecture Notes in Mathematics-969: Combinatorial Theory. Ed: D. Jungnickel, K. Vedder. Springer-Verlag, 1982, 326 pp, $16 (P). [ISBN: 0-387-11971-X] A collection of 21 research papers on a variety of topics in combinatorics that constitute the proceedings of the Conference on Combinatorial theory at Schloss Rauischholzhausen in May 1982, marking the 375th anniversary of Universität Giessen. SS

Combinatorics, P, L. Combinatorics on Words. M. Lothaire. Encyclopedia of Math. & Its Applic., V. 17. Addison-Wesley, 1983, xix + 238 pp, $32.95. [ISBN: 0-201-13516-7] A collaborative collection of articles which thoroughly and nicely survey and unify much recent (and older) research on "words," i.e., strings of symbols, as both an algebraic and combinatorial subject. Exposition is sufficiently elementary for advanced undergraduates, computer scientists and linguists. Many problems, extensive bibliography. GHM

Number Theory, S(13), P, L. The Lore of Prime Numbers. George P. Loweke. Vantage Pr, 1982, viii + 259 pp, $17.95. [ISBN: 533-04347-6] A history of number theory which focuses on the study of primes up until 1966. The approach is expository so it includes no difficult proofs. Interesting reading.

Well documented. CEC

Number Theory, S(18), P. Number Theory Related to Fermat's Last Theorem: Proceedings of the Conference Sponsored by the Vaughn Foundation. Ed: Neal Koblitz. Progress in Math., V. 26. Birkhauser Boston, 1982, x + 362 pp, $30. [ISBN: 3-7643-3104-6] This volume is the result of a meeting at MIT's Endicott House in September, 1981 on current mathematical work relating to Fermat's last theorem. CEC

Linear Algebra, S(16-17), L. Hankel and Toeplitz Matrices and Forms: Algebraic Theory. I.S. Iohvidov. Transl: G. Philip A. Thijsse. Birkhauser Boston, 1982, xiii + 231 pp, $24.95. [ISBN: 3-7643-3090-2] Translation of 1974 Russian text, with up-dated references; exercises. Hankel (resp. Toeplitz) matrices are those whose entries depend only on the sum (difference) of the indices. This purely algebraic study is based on classical linear algebra; inertia, signature, etc., are introduced and applied. RB

Linear Algebra, T(14-17: 1, 2), S, P, L. Matrices with Applications in Statistics, Second Edition. Franklin A. Graybill. Wadsworth Pub, 1983, xii + 461 pp, $31.95. [ISBN: 0-534-98038-4] Presupposes a course in matrix or linear algebra. Considers topics useful in connection with multivariate analysis and the general linear model, including generalized inverses, patterned matrices, integration and differentiation, inverse positive matrices, projections. FLW

Linear Algebra, T(14-15), S. An Introduction to Linear Algebra. L. Mirsky. Dover Pub, 1982, xi + 433 pp, $8 (P). [ISBN: 0-486-61547-2] Reprint of 1955 text. A classical approach near the beginning is gradually supplanted by a more abstract viewpoint; concreteness throughout. Filled with historical references. Some dated or non-standard terminology. Numerous exercises extend the presentation, and there are many miscellaneous problems. RB

Linear Algebra, T(14-15: 1). Applied Linear Algebra with APL. Garry Helzer. Little, Brown & Co, 1983, xv + 588 pp, $35 [ISBN: 0-316-35526-7]; Instructor's Manual, v + 52 pp, (P). An introduction to linear algebra emphasizing those ideas of greatest importance in applications using APL as a teaching tool. Not as much abstraction as many texts, but more computational algorithms and applications than most. AO

Category Theory, P. Lecture Notes in Mathematics-962: Category Theory. Ed: K.H. Kamps, D. Pumplün, W. Tholen. Springer-Verlag, 1982, xv + 322 pp, $16 (P). [ISBN: 0-387-11961-2] These proceedings of the international conference held at Gummersbach, July 6-10, 1981 contain applications to algebra, logic and topology. JAS

Algebra, T*(17-18: 1, 2), P. Associative Algebras. Richard S. Pierce. Grad. Texts in Math., V. 88. Springer-Verlag, 1982, xii + 436 pp, $39. [ISBN: 0-387-90693-2] Designed to bridge from early graduate study to the beginning of thesis specialization, starting with Wedderburn theory. After structure and representation theorems, second half concerns central simple algebras (studied via Brauer groups of fields). Well organized: short, modular chapters and sections, effective symbol indexing, sizable bibliography, many exercises. RB

Algebra, P. Proceedings of Conference on Near-rings and Near-fields. Ed: G. Ferrero, C. Ferrero Cotti. Universita degli Studi di Parma, 1981, vii + 224 pp, (P). Lectures, papers, abstracts from the week-long conference held in San Benedetto del Tronto (Italy) in 1981. LCL

Algebra, S(18), P. Lecture Notes in Mathematics-958: Group Extensions, Representations, and the Schur Multiplicator. F. Rudolf Beyl, Jürgen Tappe. Springer-Verlag, 1982, iv + 278 pp, $13.50 (P). [ISBN: 0-387-11954-X] A unified treatment of several areas of group theory, using the Schur multiplicator as the key concept and interpreting it in the light of modern homology and cohomology theory. Applications to projective representations, isoclinism, and assorted related topics. Indexes, bibliography. JS

Algebra, T(14-16: 1, 2), S. Fundamental Concepts of Algebra. Bruce E. Meserve. Dover Pub, 1981, ix + 294 pp, $6 (P). [ISBN: 0-486-61470-0] An unabridged and corrected republication of the second printing (1959) of the work originally published in 1953. This pre-"modern-math" text places much emphasis on the theory of polynomials and the theory of equations. Other chapters include number theory, determinants and matrices, geometrical constructions, and curve sketching. LCL

Algebra, T(16-18), S, P. A Survey on Congruence Lattice Representations. E. Tamás Schmidt. Teubner-Texte, Band 42. BG Teubner, 1982, 115 pp, $5 (P). A survey of the work that has been done on the problem of representing a lattice as the congruence lattice of some special type of algebra. Exercises, bibliography. LCL

Algebra, P. Seminar D. Eisenbud/ B. Singh/ W. Vogel/ Volume 2. D. Eisenbud, B. Singh, W. Vogel. Teubner-Texte, Band 48. B.G. Teubner, 1982, 108 pp, 14M (P). Eight papers on commutative rings and algebraic geometry. JD-B

Finite Mathematics, T(13: 1). College Mathematics with Business Applications, Third Edition. John E. Freund, Thomas A. Williams. Prentice-Hall, 1983, xvii + 445 pp, $24.95. [ISBN: 0-13-146498-1] Changes from the Second Edition include: more intuitive treatment of limits; review of elementary algebra; new chapter on mathematics of finance; chapters on simulation and periodic functions have been dropped. New end-of-chapter summaries, glossaries, and supplementary exercises. (First

Edition, TR, August-September 1969; Extended Review, January 1971; Second Edition, TR, November 1975.) JRG

Finite Mathematics, T(13-15: 1). Applied Finite Mathematics for the Managerial and Social Sciences. Soo Tang Tan. Prindle, Weber & Schmidt, 1983, ix + 470 pp. [ISBN: 0-87150-336-0] Conventional in content and straightforward in presentation, elementary treatment of linear programming (including simplex method), probability and statistics, Markov chains and game theory, mathematics of finance. Ample, largely routine exercises with answers, tables, index. JS

Calculus, T(14-16: 1, 2), S, L. Advanced Calculus, Third Edition. Angus E. Taylor, W. Robert Mann. Wiley, 1983, xv + 732 pp, $31.95. [ISBN: 0-471-02566-6] A complete and comprehensive text at the intermediate-to-advanced undergraduate level. Covers a good deal of analysis, e.g., uniform convergence, infinite series, mean-value theorems. Uses and assumes linear but not multilinear algebra. Extensive explanation accompanies the theoretical development. Some calculator-aided numerical problems (Second Edition, TR, February 1973). PZ

Calculus, T(13: 1, 2), S. Calculus: Pure and Applied. A.J. Sherlock, E.M. Roebuck, M.G. Godfrey. Edward Arnold Pub, 1982, x + 534 pp, $16.95 (P). [ISBN: 0-7131-3446-1] Polynomial calculus (the first quarter) followed by algebraic and transcendental functions, with applications interspersed. Nontheoretical, computational, special emphasis on numerical methods and differential equations (calculus as a language for science). LCL

Calculus, T(13-14: 3). Calculus and Analytic Geometry. J. Douglas and Barbara Trader Faires. Prindle, Weber & Schmidt, 1982, 11 + 1141 pp. [ISBN: 0-87150-323-9] Stand-out features include numerous historical notes, over 6000 exercises, many with applications, and about fifty problems (but not solutions) from Putnam competitions. Trigonometric functions are introduced early. Readable, attractive two-color format, flexible. Complete solutions manual and two-volume student supplement are available. JK

Differential Equations, P. Instability, Nonexistence and Weighted Energy Methods in Fluid Dynamics and Related Theories. B. Straughan. Research Notes in Math., No. 74. Pitman Pub, 1982, 169 pp, $18.95 (P). [ISBN: 0-273-08564-6] The author illustrates several techniques which lead to nonexistence results for solutions to partial differential equations; he also presents variations of a weighted energy method which has been used to obtain results for a variety of problems in mathematical physics. LCL

Differential Equations, P. Lecture Notes in Mathematics-957: Differential Equations. Ed: D.G. de Figueiredo, C.S. Hönig. Springer-Verlag, 1982, viii + 301 pp, $16 (P). [ISBN: 0-387-11951-5] Proceedings of the First Latin American School of Differential Equations held at São Paulo, Brazil, June 29 to July 17, 1981. JAS

Differential Equations, P. The Fundamental Principle for Systems of Convolution Equations. Daniele Carlo Struppa. Memoirs No. 273. AMS, 1983, iv + 167 pp, $10 (P). [ISBN: 0-8218-2273-X] Exposition of an extension of Ehrenpreis' "fundamental principle" in the theory of systems of linear partial differential equations. Extends methods of Berenstein and Taylor to establish admission of Fourier representation of solutions in certain spaces. Brief discussion of some open problems. Good bibliography. JK

Differential Equations, T(15-17: 1), S, L*. Topics in Ordinary Differential Equations. William D. Lakin, David A. Sanchez. Dover Pub, 1982, vi + 154 pp, $4 (P). [ISBN: 0-486-61606-1] Unabridged and slightly corrected republication of 1970 edition (TR, February 1971). Informal, intuitive approach. Little theory. Few proofs. Perturbation methods link five chapters. Prerequisites include introductory course in differential equations but not in complex variables. Not intended to be exhaustive treatment. Problems. References. JK

Differential Equations, P. Generic Bifurcations for Involutory Area Preserving Maps. Russell J. Rimmer. Memoirs No. 272. AMS, 1983, v + 165 pp, $10 (P). The use of involutory area preserving maps in studying generic bifurcations associated with symmetric periodic solutions of Hamiltonian systems. JG

Differential Equations, P. The Stability of Multi-dimensional Shock Fronts. Andrew Majda. Memoirs No. 275. AMS, 1983, iv + 95 pp, $6 (P). A systematic study of the linearized stability of multidimensional shock-fronts, with special emphasis on applying the theory to shock-fronts for the physical equations of gas dynamics. LCL

Differential Equations, S(18), P. Quasikonforme Abbildungen und elliptische Systeme erster Ordnung in der Ebene. Heinrich Renelt. Teubner-Texte, Band 46. B.G. Teubner, 1982, 140 pp, 14,50M (P). On systems of first order elliptic partial differential equations in the plane whose solutions are composites of analytic functions and quasiconformal mappings (Bers-Nirenberg). JD-B

Differential Equations, S(17-18), P. Lectures on Exponential Decay of Solutions of Second-Order Elliptic Equations: Bounds on Eigenfunctions of N-Body Schrödinger Operators. Shmuel Agmon. Math. Notes, No. 29. Princeton U Pr, 1982, 118 pp, $10.50 (P). A study of estimates for solutions of second order elliptic equations in unbounded domains. JG

Numerical Analysis, P. Analysis of Global Expansion Methods: Weakly Asymptotically Diagonal Systems. L.M. Delves, T.L. Freeman. Comp. Math. & Applic. Academic Pr, 1981, x + 275 pp, $49. [ISBN: 0-12-208880-8] WAD systems, used in solving differential and integral equations, concern the convergence of expansions of f(x) into sums of terms of the form $a^N{}_i h_i(x)$ for i = 1,2,...,N. Some new results, an account of known but scattered results, and an emphasis on computability and error analysis. AWR

Numerical Analysis, P. Lecture Notes in Mathematics-968: Numerical Integration of Differential Equations and Large Linear Systems. Ed: Juergen Hinze. Springer-Verlag, 1982, vi + 412 pp, $20 (P). [ISBN: 0-387-11970-1] Proceedings of two workshops held at the University of Bielefeld in the spring of 1980. JAS

Numerical Analysis, P. Lecture Notes in Mathematics-965: Topics in Numerical Analysis. Ed: P.R. Turner. Springer-Verlag, 1982, ix + 202 pp, $11 (P). [ISBN: 0-387-11967-1] Proceedings of a 1981 summer school at the University of Lancaster: lecture notes from five of the courses, covering Volterra and Abel integral equations, multivariate approximation methods, splines, finite element methods, and piecewise linear homotopy algorithms. LAS

Numerical Analysis, P. Lecture Notes in Mathematics-960: Multigrid Methods. Ed: W. Hackbusch, U. Trottenberg. Springer-Verlag, 1982, vii + 652 pp, $28.80 (P). [ISBN: 0-387-11955-8] Proceedings of a November 1981 conference held at Köln-Porz: three systematic introductory papers (with a uniform notation for essential concepts) followed by 18 specific papers on theory, applications and software developments. LAS

Numerical Analysis, P. Tables for Numerical Integration. H.V. Smith. Macmillan Pub, 1982, vi + 26 pp, $5.95 (P). [ISBN: 0-02-852670-8] Tables of global error bounds for many of the well-known quadrature formulae; a short introduction includes examples to illustrate their use. LCL

Numerical Analysis, P. Birkhoff Interpolation. G.G. Lorentz, K. Jetter, S.D. Riemenschneider. Ency. of Math. & Its Applic., V. 19. Addison-Wesley, 1983, lv + 237 pp, $32.50. [ISBN: 0-201-13518-3] Polynomial interpolation given irregular information about the values of the functions and its derivatives. Applications to approximation, quadrature, splines and Chebyshev systems. Uses interpolation matrices, coalescence of knots, independent knots among other techniques. RWN

Functional Analysis, S(18), P**, L.** A Hilbert Space Problem Book, Second Edition, Revised and Enlarged. Paul R. Halmos. Grad. Texts in Math., No. 19. Springer-Verlag, 1982, xvii + 369 pp, $28. [ISBN: 0-387-90685-1] A Second Edition of the justly famous original. 250 problems, 60 of them new since the First Edition (TR, April 1968), with remarks, hints, and complete solutions. The choice of interesting problems and their presentation is outstanding; the author's remarks on (and beyond) the problems are invaluable. PZ

Functional Analysis, P. Invariant Subspaces and Other Topics. Ed: C. Apostol, et al. Operator Theory: Advances & Applic., V. 6. Birkhauser Boston, 1982, 229 pp, $26.95. [ISBN: 3-7643-1360-9] The proceedings of the Sixth International Conference on Operator Theory held in Timisoara and Herculane (Romania), June 1-11, 1981. This volume contains eighteen papers contributed by the conference participants on invariant subspaces as well as a wide range of other topics (e.g., operator algebras, dilation theory). AO

Functional Analysis, P. Spectral Methods in Linear Transport Theory. H.G. Kaper, C.G. Lekkerkerker, J. Hejtmanek. Operator Theory: Advances & Applic., V. 5. Birkhauser Boston, 1982, xi + 345 pp, $29.95. [ISBN: 3-7643-1372-2] This monograph provides a mathematical theory for linear transport processes in matter (e.g., neutron transport in nuclear reactors, photon transport in planetary and stellar atmosphere). The interest is on functional-analytic and spectral-theoretic methods. This branch of applied mathematics provides several nontrivial applications of functional analysis: long live Hilbert space! LCL

Algebraic Geometry, P. Lecture Notes in Mathematics-956: Group Actions and Vector Fields. Ed: J.B. Carrell. Springer-Verlag, 1982, 144 pp, $8 (P). [ISBN: 0-387-11946-9] Proceedings (eleven research papers) of the seminar on group actions and vector fields held at the University of British Columbia in January-February 1981. LCL

Algebraic Geometry. Lecture Notes in Mathematics-959: Géométrie Algébrique Réelle et Formes Quadratiques. Ed: J.-L. Colliot-Thélène, et al. Springer-Verlag, 1982, x + 458 pp, $22 (P). [ISBN: 0-387-11959-0] Proceedings of a conference held at the Université de Rennes in May 1981. JAS

Algebraic Geometry, T(18: 1), S, P. Vorlesungen über algebraische Flächen. Herbert Kurke. Teubner-Texte, Band 43. BG Teubner, 1982, 204 pp, $7.92 (P). A monograph on the modern theory of algebraic surfaces. JD-B

Algebraic Geometry, P. Curves in Projective Space. Joe Harris, David Eisenbud. Presses U Montreal, 1982, 138 pp, $13 (P). [ISBN: 2-7606-0603-1] Considers the problem of describing curves in projective space, emphasizing two methods of parametrizing projective curves. Chow and Hilbert schemes are discussed, along with two other classificatory approaches the Brill-Noether theory and the theory Castelnuovo theory. SG

Algebraic Geometry, P. Enumerative Geometry and Classical Algebraic Geometry. Ed: Patrick Le Barz, Yves Hervier. Progress in Math., V. 24. Birkhauser Boston, 1982, x + 252 pp, $20. [ISBN: 3-7643-3106-2] Partial proceeding of a conference held at Nice in July, 1981. The papers deal with low-dimensional projective spaces and curves over algebraically closed fields. SG

Differential Geometry, S(17-18), P, L. Curvature and Homology. Samuel I. Goldberg. Dover Pub, 1982, xvii + 315 pp, $6.50 (P). [ISBN: 0-486-64314-X] An unabridged and corrected republication of the 1962 edition. A study of the curvature and homology of Riemannian manifolds, as well as Kaehler manifolds. Topics include Betti numbers, Lie groups, and groups of transformations. JG

Differential Geometry, T(17), S, L. Calcul différentiel et géométrie. Daniel Leborgne. Presses U France, 1982, 262 pp, (P). [ISBN: 2-13-037495-6] A modern approach to differential geometry from the viewpoint of differentiable manifolds, tangent spaces and differential forms. JG

Differential Geometry, T(17-18: 1), S, P, L. Lectures on Differential Geometry, Second Edition. Shlomo Sternberg. Chelsea Pub, 1983, xv + 442 pp, $25. [ISBN: 0-8284-0316-3] Apart from correction of typographical errors, the text is unchanged from the first edition but enhanced by two appendices: Appendix III is a complete reproduction of a paper co-authored with Guillemin on transitive differential geometry, and Appendix IV is a brief survey by the author of relevant developments of the past twenty years together with references. JS

Geometry, S(17-18), P. Finite Geometries: Proceedings of a Conference in Honor of T.G. Ostrom. Ed: Norman L. Johnson, Michael J. Kallaher, Calvin T. Long. Lect. Notes in Pure & Appl. Math., V. 82. Dekker, 1983, xii + 454 pp, $55 (P). [ISBN: 0-8247-1052-5] An extremely interesting festschrift consisting of 36 papers on finite geometries, related algebraic and combinatorial topics, and their histories. SS

Geometry, S(13), L. Famous Problems of Geometry and How to Solve Them. Benjamin Bold. Dover Pub, 1982, xii + 112 pp, $3 (P). [ISBN: 0-486-24297-8] An unabridged and slightly corrected version of the work originally published in 1969 as Famous Problems of Mathematics: A History of Construction with Straight Edge and Compass. Includes several famous classical construction problems and the analytic criteria which demonstrate the impossibility of their construction via straight edge and compass. JNC

Topology, T(15-17: 1, 2).** Differential Topology: An Introduction. David B. Gauld. Pure & Appl. Math., V. 72. Dekker, 1982, v + 241 pp, $29.75. [ISBN: 0-8247-1709-0] Neither pure point set topology nor technical and formal differential topology, but rather geometric topology with differentiation as a tool and as a link to the geometric environment of multivariable analysis. This is a well written text with a number of exercises. There is a helpful index of notation, but the main index is rather brief. JAS

Topology, P. Beta-homotopy Equivalences Have Alpha-cross Sections. Luis Montejano. Memoirs No. 274. AMS, 1983, iv + 37 pp, $5 (P).

Operations Research, T(15-17), S. The High Cost of Clean Water: Models for Water Quality Management. Edward Beltrami. UMAP Expos. Mono. Ser. Birkhauser Boston, 1982, 53 pp, $8.95 (P). [ISBN: 3-7643-3098-8] A prototypical water-pollution case study based on similar studies done on a more extensive scale in the Long Island area of New York. Several mathematical models (e.g., using differential equations, integer programming) are presented to help articulate and analyze the trade-offs between costs and water quality. The complexities of the real-world problem are discussed and care is taken in noting simplifying assumptions in the models. Includes exercises with solutions, as well as suggestions for further research. LCL

Optimization, P. Bonn Workshop on Combinatorial Optimization. Ed: Achim Bachem, Martin Grötschel, Bernard Korte. Math. Stud., No. 66. Elsevier North-Holland, 1982, ix + 312 pp, $55.75 (P). [ISBN: 0-444-86366-4] Eighteen papers covering a broad spectrum of optimization problems which arise in graph theory (matching, independence, paths). LCL

Probability, S*(15-17), L*. Probability, Statistics and Truth, Second Revised English Edition. Richard von Mises. Dover Pub, 1981, xii + 244 pp, $4 (P). [ISBN: 0-486-24214-5] Reprint of the 1957 English edition of this classic, which was a translation of the definitive 1951 German Third Edition. A pioneer effort in the foundations of probability theory. RSK

Probability, T(16-17: 1, 2), S, P, L. Probability and Random Processes. Geoffrey Grimmett, David Stirzaker. Clarendon Pr, 1982, ix + 354 pp, $45; $19.95 (P). [ISBN: 0-19-853184-2; 0-19-853185-0] A rigorous introduction to probability theory with minimal reference to measure theory, followed by individual chapters on random processes: Markov chains, stationarity and diffusion, renewals, queues. A nice balance between formal and informal, theory and application. LCL

Probability, T(18: 1), S, P*. Brownian Motion and Diffusion. David Freedman. Springer-Verlag, 1983, xii + 231 pp, $24. [ISBN: 0-387-90805-6] Originally published by Holden-Day (1971) (TR, October 1971). Brownian motion as a probability on a Borel σ-field. Diffusion is obtained from Brownian motion by changing time. Emphasizes variation, reflection and invariance principles. Includes examples, heuristic arguments, motivation as well as formal proofs. SG

Statistics, T(15-17: 1, 2), S. Introductory Engineering Statistics, Third Edition. Irwin Guttman, S.S. Wilks, J. Stuart Hunter. Wiley, 1982, xi + 580 pp, $33.95. [ISBN: 0-471-07859-X] This edition includes new material on reliability and contrasts, revisions of other topics, and many new problems and data sets (Second Edition, TR, April 1972). FLW

Statistics, S(17-18), P. Theory of Statistical Experiments. H. Heyer. Ser. in Stat. Springer-Verlag, 1982, x + 289 pp, $19.80. [ISBN: 0-387-90785-8] A functional analytic treatment of statistical decision theory. FLW

Statistics, P. Higher Order Asymptotics for Simple Linear Rank Statistics. R.J.M.M. Does. Math. Centre Tracts, No. 151. Math Centrum, 1982, 91 pp, Dfl. 11,55 (P). [ISBN: 90-6196-243-9] Technical monograph dealing with the problem of obtaining Berry-Esseen bounds and Edgeworth expansions for simple linear rank statistics. RSK

Statistics, P. Lecture Notes in Statistics-13: Contributions to a General Asymptotic Statistical Theory. J. Pfanzagl, W. Wefelmeyer. Springer-Verlag, 1982, vii + 315 pp, $16.80 (P). [ISBN: 0-387-90776-9] Presents the beginnings of a unified asymptotic theory, covering both parametric and nonparametric models, but restricted to normal approximations. Good set of references. RSK

Statistics, P. Time Series Analysis: Theory and Practice 1. Ed: O.D. Anderson. Elsevier North-Holland, 1982, ix + 756 pp, $93. [ISBN: 0-444-86337-0] Proceedings of the International Congress held at Valencia, Spain, June 1981. Contains 48 papers, in addition to an introduction by the editor. RSK

Statistics, P. Kernel Discriminant Analysis. D.J. Hand. Electronic & Electrical Engin. Res. Stud., V. 2. Research Studies Pr (US Distr: Wiley), 1982, x + 253 pp, $37.95. [ISBN: 0-471-10211-3] Provides a general nonparametric method for attacking problems of "classification" (in the areas of discriminant analysis and pattern recognition rather than cluster analysis). Includes a good set of references to this relatively new technique. RSK

Statistics, T*(14: 1, 2). Modern Business Statistics. Ronald L. Iman, W.J. Conover. Ser. in Prob. & Math. Stat. Wiley, 1983, xxxiv + 777 pp, $27.95 [ISBN: 0-471-09668-7]; Introduction to Modern Business Statistics. W.J. Conover, Ronald L. Iman. Ser. in Prob. & Math. Stat. Wiley, 1983, xxxi + 525 pp, $25.95. [ISBN: 0-471-09669-5] Introduction is the first thirteen chapters of Modern Business Statistics. Both contain contingency tables, correlation, regression, and time series analysis, in addition to other standard topics. Modern Business Statistics also contains multiple regression, analysis of variance and covariance, and decision theory. Novel features include the emphasis on assumptions, tests of these assumptions (particularly normality), and parallel presentations of nonparametric alternatives. RSK

Statistics, T(13-14: 1). Elementary Statistics, Second Edition. Mario F. Triola. Benjamin/Cummings Pub, 1983, xvi + 496 pp, $21.95. [ISBN: 0-8053-9320-X] Presupposes only high school algebra. Covers the usual topics, including some nonparametric methods. Many examples use real data. Optional computer projects included. FLW

Statistics, S(15-17), P. Applied Statistics: A Handbook of Techniques. Lothar Sachs. Transl: Zenon Reynarowych. Series in Stat. Springer-Verlag, 1982, xxviii + 706 pp, $48. [ISBN: 0-387-90558-8] Translation and expansion of the 1978 German Fifth Edition. Written for non-mathematicians, it has as its main concern practical applications. Particular emphasis is given to small sample problems and distribution free methods. Extensive bibliography. RSK

Statistics, P. A Festschrift for Erich L. Lehmann: In Honor of His Sixty-Fifth Birthday. Ed: Peter J. Bickel, Kjell A. Doksum, J.L. Hodges, Jr. Statistics/Probability Ser. Wadsworth Pub, 1983, viii + 461 pp, $39.95. [ISBN: 0-534-98044-9] Twenty-eight statistical papers written by colleagues, students, and friends of Erich Lehmann. Concludes with a bibliography of Lehmann's works. RSK

Statistics, T*(15-17: 1, 2), L. Applied Linear Regression Models. John Neter, William Wasserman, Michael H. Kutner. Richard D. Irwin, 1983, xv + 547 pp, $27.50. [ISBN: 0-256-02547-9] Revision of the regression portion of Neter and Wasserman's 1974 text Applied Linear Statistical Models (TR, December 1974). New topics include detection of multi-collinearity, ridge regression, detection of influential observations, and nonlinear regression. Examples from the health and biological sciences have been added, and problem sets have been substantially expanded. RSK

Statistics, S(17-18), P. Grundlagen der sequentiellen Statistik. Hartmut Heckendorff. Teubner-Texte, Band 45. BG Teubner, 1982, 166 pp, $7.50 (P). A sophisticated introduction to the theory of sequential statistical experiments. Contains a good deal on Bayesian decision theory. JD-B

Statistics, P. Pao-Lu Hsu Collected Papers. Ed: Kai Lai Chung. Springer-Verlag, 1983, xii + 589 pp, $48. [ISBN: 0-387-90725-4] Contains reprints (in English) of all of Hsu's published mathematical papers. Also includes biographical information and short discussions of Hsu's work in inference, multivariate analysis, and probability. RSK

Statistics, T(17: 1), P. Multivariate Statistical Methods, An Introduction. Marvin J. Karson. Iowa St U Pr, 1982, x + 307 pp, $24.95. [ISBN: 0-8138-1845-1] Assumes background in multivariate calculus, matrix algebra, and mathematical statistics, which are briefly reviewed. Includes chapters on correlations, discriminant analysis and allocation, principal components analysis,

factor analysis, cluster analysis, and discrete multivariate analysis, in addition to the usual introductory material. Contains many numerical examples. RSK

Statistics, T(14-18: 1), S, P. APL-STAT: A Do-It-Yourself Guide to Computational Statistics Using APL. James B. Ramsey, Gerald L. Musgrave. Lifetime Learning Pub, 1981, xi + 340 pp, $17.95 (P). [ISBN: 0-534-97985-8] Explains how to use APL to carry out statistical calculations. Presupposes knowledge of statistics. Not an introduction to APL but the authors say many have learned APL from it. FLW

Statistics, S(15-18)), P, L. The Human Side of Statistical Consulting. James R. Boen, Douglas A. Zahn. Lifetime Learning Pub, 1982, xii + 196 pp, $15.95. [ISBN: 0-534-97949-1] This helpful and interesting book deals with the "human, behavioral, and political issues that a consultant confronts in organizations and in dealing with people." FLW

Statistics, T(15-18: 1, 2), S, P, L. Statistical Methods for Survival Data Analysis. Elisa T. Lee. Lifetime Learning Pub, 1980, xiii + 557 pp, $31.95. [ISBN: 0-534-97987-4] Parametric and non-parametric methods for survival data analysis. Many examples and exercises use real medical data. 46 pages of Fortran program listings. 65 pages of tables. Extensive bibliography. FLW

Computer Literacy, T?(12), L? Armchair BASIC: An Absolute Beginner's Guide to Programming in BASIC. Annie and David Fox. Osborne/McGraw-Hill, 1983, vii + 264 pp, $11.95 (P). [ISBN: 0-931988-92-6] Intended to explain popular computer terminology and Basic to "the casually curious as well as the utterly baffled," without the need for a computer. No mention of program development, program structure; short giveaway quizzes, no exercises; few program examples exceed 10 lines. Glossary. RB

Computer Literacy, S, L. Apple II User's Guide. Lon Poole, Martin McNiff, Steven Cook. Osborne/McGraw-Hill, 1981, xii + 385 pp (P). [ISBN: 0-931988-46-2] For Apple II and Apple II+, concerning everything from locating the "on" switch to discussions of graphics, machine language monitor program, etc. Tutorials introduce Basic from a hardware viewpoint--not suitable for most introductory courses. A complete description of all Basic commands is included. An unusually accessible technical reference. RB

Computer Literacy, T*(13: 1). Introduction to Computers and Data Processing. Gary B. Shelly, Thomas J. Cashman. Anaheim Pub, 1980, (P). xvi + 458 pp [ISBN: 0-88236-115-5]; Student Workbook and Study Guide, 247 pp [ISBN: 0-88236-116-3]; Answer Manual for Student Workbook and Study Guide, 191 pp; Transparency Masters, 187 pp; Test Bank, 188 pp; Instructor's Guide and Answer Manual, 173 pp. Well organized with many color photographs. An introduction to data processing that emphasizes state-of-the-art technology and techniques. AO

Computer Programming, S(13-18). Programming in Modula-2. Niklaus Wirth. Springer-Verlag, 1982, 176 pp, $13.50. [ISBN: 0-387-11674-5] Intended for someone who already knows how to program; there is, however, an introduction for the beginner. Modula-2 is a direct descendant of Pascal and Modula. The notation of the language itself encourages the principles of structured programming. Language includes all aspects of Pascal plus the module and multiprogramming concepts. Text includes the Report, two appendices, and an index. RJA

Computer Programming, S??. Your IBM Personal Computer: Use, Applications, and BASIC. David E. Cortesi. Holt, Rinehart & Winston, 1982, xiv + 253 pp, $16.95 (P). [ISBN: 0-03-061979-3] An appalling book at the junior-high-school level. The style is condescending and the content excessively fluffy--even for beginners. For example, "The prefix kilo, which usually means 1,000, here stands for a multiple of 1,024. There is a practical but uninteresting reason for this odd usage,..." No explanation whatsoever is provided--the issue is dropped. JAS

Computer Programming, T*(13: 1). Structured Problem Solving with Pascal. Lawrence J. Mazlack. Holt, Rinehart & Winston, 1983, xiii + 386 pp, $21.95 (P). [ISBN: 0-03-060153-3] An introduction to top-down analysis and structured programming using Pascal. This book provides a relatively complete introduction to the Pascal language. AO

Computer Programming, T(13-14). Introduction to Programming with ESP and Pascal. Allen B. Tucker, Jr. Holt, Rinehart & Winston, 1983, xvi + 363 pp, $18.95 (P). [ISBN: 0-03-059148-1] ESP = Eight Statement Pascal, the subset used first in the text. Overlaps other books of the author's on Pascal and Basic. RWN

Computer Programming, T(13: 1). Programming the IBM Personal Computer: BASIC. Neill Graham. Holt, Rinehart & Winston, 1982, x + 287 pp, $16.95 (P). [ISBN: 0-03-061911-4] Integrates introductory Basic programming with the features of the IBM personal computer. RWN

Computer Programming, T(13-14: 1), S, L.** Problem Solving with BASIC. Richard Dillman. Holt, Rinehart & Winston, 1983, x + 326 pp, $18.95 (P). [ISBN: 0-03-061981-5] An introductory text, usable with micros, with solid emphasis on top-down design, standard programming structures, maintainable programs, learning to read manuals, and the value of these principles in commercial programming. Effective application of "analysis and design as a prerequisite for coding" from the beginning. Relatively limited instruction set. A well thought out, convincing book. RB

Computer Programming, S(13-15). Programming the PL/I Way. Dan Smedley. TAB Books, 1982, x + 288 pp, $9.95 (P). [ISBN: 0-8306-1414-1] An introduction to the programming language PL/I. The text assumes no prior knowledge of programming and begins with the basic aspects of the language. It concentrates almost exclusively on the syntactic/semantic aspects of the language and does not branch out into broader issues of programming in general. It would be appropriate for those interested specifically in learning the PL/I language. MS

Computer Programming, P. TRS-80 Color Programs. Tom Rugg, Phil Feldman. TAB Books, 1982, 328 pp, $25.95. [ISBN: 0-8306-1481-8] This book is simply a collection of 36 Basic programs which can be executed on the TRS-80 color computer manufactured by Radio Shack. The programs are in the area of games, mathematics, graphics, home use, and educational use. Each program is accompanied by extensive documentation about how to run it and interpret the output. MS

Computer Programming, T?(13). Programming the IBM Personal Computer: UCSD Pascal. Seymour V. Pollack. Holt, Rinehart & Winston, 1983, xi + 323 pp, $18.95 (P). [ISBN: 0-03-062637-4] Suitable for quick retooling from Fortran, Basic, etc., into Pascal syntax, but not necessarily into the spirit of the language. Type declarations are discussed early, procedures and functions are postponed. Hardware-inspired understanding of procedure and function action. Not strong on algorithms, style; cryptic identifiers. RB

Computer Programming. The PASCAL Handbook. Jacques Tiberghien. Sybex, 1981, xix + 485 pp, $19.95 (P). [ISBN: 0-89588-053-9] Each keyword, symbol, standard identifier, identifier in at least one of Standard, HP1000, CDC, OMSI, Pascal/Z, and UCSD Pascals, is discussed in alphabetical order. Each entry includes syntax, description, implementation dependent features, and an example. Most are not worth the space, but others might be helpful to programmers writing software for different machines or who are interested in knowing the various extensions for the listed compilers. RBK

Computer Programming. Introduction to the UCSD p-System. Charles W. Grant, Jon Butah. Sybex, 1982, xiv + 300 pp, $14.95 (P). [ISBN: 0-899588-061-X] A useful, well-written, well-organized reference for a p-system user. Explicit examples are given for the most common tasks in editing, filing, and compiling. Includes chapters on preparing short and large-scale Pascal programs. Examples from Version II.0, but information is valid for versions through IV. RBK

Computer Programming, T(13), S, L. Professional Programming Techniques--Starting with the BASICs. Richard Galbraith. TAB Books, 1982, vi + 301 pp, $10.95 (P). [ISBN: 0-8306-0128-7] A readable introduction to programming in Basic with emphasis on structure, style, and documentation. JG

Computer Programming, S??(13). FORTRAN Programs for Scientists and Engineers. Alan R. Miller. Sybex, 1982, xvii + 280 pp, $15.95 (P). [ISBN: 0-89588-082-2] This book contains a collection of Fortran subroutines suitable for use on a microcomputer which implement simple numerical algorithms for such problems as random number generation, the solution of systems of linear equations, least squares curve fitting, and numerical integration. The numerical algorithms presented seem to have been chosen for their simplicity rather than their accuracy. AO

Data Structures, T*(14-7: 1, 2), S. Design of Computer Data Files. Owen Hanson. Computer Software Engin. Ser. Computer Sci Pr, 1982, x + 358 pp, $24.95. [ISBN: 0-914894-17-X] Provides single source of information on factors influencing file design. Topics include backing storage devices, record formats, blocking and buffering, sequential and direct file organization, multiple-key processing. Chapter references and questions. Appendices. Index. RJA

Software Systems, T(15-17: 1), S, P?, L?. Introduction to Arithmetic for Digital Systems Designers. Shlomo Waser, Michael J. Flynn. Holt, Rinehart & Winston, 1982, xvii + 308 pp, $35.95. [ISBN: 0-03-060571-7] A poorly written assemblage of basic facts and state-of-the-art algorithms for doing machine arithmetic (fixed and floating point), stressing "information of practical implementation value." Abounds with redundancies, non sequiturs, sentence fragments and just plain awkwardness. Appendix contains specifications on many currently available MSI and LSI arithmetic chips. GHM

Computer Science, T(15-16: 1), P. Computer Network Architectures. Anton Meijer, Paul Peeters. Computer Sci Pr, 1982, xi + 396 pp, $27.95. [ISBN: 0-914894-41-2] This text reviews the design and implementation of distributed computer systems. It uses the ISO Open System Model as its frame of reference and discusses the various levels of protocol in that model. Part I discusses protocols in general, while parts II and III look at specific examples of architectures, both private (e.g., DECNET and SNA) and public (e.g., X.25). The book is intended for professionals in the field of computer networks and distributed systems, especially those involved with design and performance evaluation of alternative protocols. MS

Computer Science, T?(13). Using Computers in Mathematics. Gerald H. Elgarten, Alfred S. Posamentier, Stephen E. Moresh. Addison-Wesley, 1983, 574 pp, $21.44. [ISBN: 0-201-10450-4] Elementary Basic through application to various topics in secondary mathematics. Sample programs are followed by assignments to produce different output, yet changes required are usually as trivial as altering the wording of a question. Mathematics is reduced to formulas, with no attempt to create understanding. Programming is old-fashioned, with emphasis on flowcharts rather than structure. MW

Computer Science, T(16-18), S, P*, L*. Computer Algebra: Symbolic and Algebraic Computation. Ed: B. Buchberger, et al. Computing Supplementum, No. 4. Springer-Verlag, 1982, vii + 283 pp, $44 (P). [ISBN: 0-387-81684-4] This book of sixteen survey articles (with extensive references) is the

first attempt at drawing together the disparate results on efficiency and implementation of algebraic algorithms scattered throughout the literature. The articles are arranged in a "top-down" manner: starting with high-level problems whose motivation is immediately apparent (e.g., algebraic simplification), it proceeds to more basic problems whose solution is needed for the problems at the higher levels (e.g., factorization of polynomials, computing with groups, computing in extension fields, computing by homomorphic images). The material in this volume can be used to supplement a traditional algebra course or as a basic reference for a course in computer algebra. LCL

Computer Science, T(14: 1). The Minicomputer in the Laboratory: With Examples Using the PDP-11, Second Edition. James W. Cooper. Wiley, 1983, xviii + 381 pp, $29. [ISBN: 0-471-09012-3] An introductory textbook in computer organization, using the PDP-11 family for all of its examples. It first introduces the PDP-11 computer system in general, and then concentrates on specific laboratory projects which can utilize the capabilities of this machine. Some of the sample laboratory projects would include plotting lab data, averaging signals, performing Fourier transforms, and detecting signal peaks. MS

Applications, P. Computerised Braille Production: Today and Tomorrow. Ed: D.W. Croisdale, H. Kamp, H. Werner. Springer-Verlag, 1983, xii + 422 pp, $17.20 (P). [ISBN: 0-387-12057-2] Proceedings of a conference held in London, May 30 to June 1, 1979. JAS

Applications (Artificial Intelligence), S(16-18), P, L. The Handbook of Artificial Intelligence, Volume II. Ed: Avron Barr, Edward A. Feigenbaum. William Kaufmann, 1982, xiii + 428 pp, $35. [ISBN: 0-86576-006-3] This volume covers three subfields of artificial intelligence: (i) artificial intelligence programming languages; (ii) expert systems in science, medicine, and education; (iii) automatic programming. List of contributors. Each chapter contains several short articles. Chapter overviews. Extensive volume bibliography. Name and subject indexes for Volume II. RJA

Applications (Biology), P. Lecture Notes in Biomathematics-47: Stochastic Transport Processes in Discrete Biological Systems. Eckart Frehland. Springer-Verlag, 1982, viii + 169 pp, $11 (P). [ISBN: 0-387-11964-7] "The subject of this volume is not to introduce the mathematics of stochastic processes but to present a field of theoretical biophysics [ion transport through biological membranes] in which stochastic methods are important." RB

Applications (Computer Organization), S(13-18), P. System Architecture. John Zarrella. Microprocessor Software Engin. Concepts Ser. Microcomputer Applic, 1980, vii + 231 pp, $10.95 (P). [ISBN: 0-935230-02-5] Intended as an introduction to advanced computer design. Integrates ideas from software engineering and computer architecture. Topics include data representations, elements, and structures; addressing and hardware communication; parallel processing, resource protection, I/O, microprogramming, error detection. Important terms are in boldface on the page where first used. Good diagrams. Glossary. References. Index. RJA

Applications (Engineering), T(16: 2), P. Signals and Systems. Alan V. Oppenheim, Alan S. Willsky, Ian T. Young. Prentice-Hall, 1983, xix + 796 pp, $32.95. [ISBN: 0-13-809731-3] Designed as an advanced undergraduate textbook in the areas of signals and electrical systems. Covers the topics of continuous time signals, discrete time signals, filtering, modulation, and both Fourier and Laplace transformations. It would be appropriate for a required one or two semester course for majors in the field of electrical engineering. Assumes a thorough grounding in calculus. MS

Applications (Engineering), P. Tutorials in Modern Communications. Ed: Victor B. Lawrence, Joseph L. LoCicero, Laurence B. Milstein. Computer Sci Pr, 1983, viii + 348 pp, $33.95. [ISBN: 0-914894-48-X] A collection of thirty-seven tutorial articles originally published in the IEEE Communications Society Magazine in the following areas: quantization and switching, data and modulation techniques, computer communications, transmission, signal processing, and secure communications. AO

Applications (Engineering), P. Annual Review of Fluid Mechanics, Volume 15, 1983. Ed: Milton van Dyke, J.V. Wehausen, John L. Lumley. Annual Reviews, 1983, 534 pp, $28. [ISBN: 0-8243-0715-1] 18 expositions of current research and mathematical models of fluid dynamics, ranging from green plants to airfoils. Begins with a survey of Ernst Mach's contributions to fluid mechanics, including numerous plates of experimental artifacts from Mach's research. LAS

Applications (Meteorology), P. Dynamic Meteorology: Data Assimilation Methods. Ed: Lennart Bengtsson, Michael Ghil, Erland Källén. Appl. Math. Sci., V. 36. Springer-Verlag, 1981, 330 pp, $18 (P). [ISBN: 0-387-90632-0] Selected lectures from the 1980 Seminar of the European Centre for Medium Range Weather Forecasts. Highly applied in nature. AWR

Applications (Modelling), S(17-18), P. Lecture Notes in Biomathematics-46: Identifiability of State Space Models. Eric Walter. Springer-Verlag, 1982, viii + 202 pp, $13.50 (P). [ISBN: 0-387-11590-0] A monograph devoted to tests of identifiability as a necessary step in the modeling process, with special attention given to transformation systems (a generalization of compartmental models). LCL

Applications (Pattern Recognition), T(17-18: 1), P. Fuzzy Techniques in Pattern Recognition. Abraham Kandel. Wiley, 1982, x + 356 pp, $34.95. [ISBN: 0-471-09136-7] Exposition of an approach to the problem of pattern classification based on the theory of fuzzy relations, fuzzy expectation, inexact matrices, et cetera. Comprehensive bibliography (3064 entries). GHM

Applications (Physics), P*. Quantum Mechanics of Large Systems. Walter Thirring. Course in Math. Physics, No. 4. Transl: Evans M. Harrell. Springer-Verlag, 1983, x + 290 pp, $32. [ISBN: 0-387-81701-8] A rigorous exposition of quantum statistical mechanics using modern techniques (e.g., C*-algebras, etc.) which summarizes much of the work in this area done in the last twenty years. A valuable addition to the mathematical physics literature. AO

Applications (Physics), P. Ray Methods for Waves in Elastic Solids: With Applications to Scattering by Cracks. J.D. Achenbach, A.K. Gautesen, H. McMaken. Mono. & Stud. in Math., No. 14. Pitman Pub, 1982, xi + 251 pp, $49.95. [ISBN: 0-273-08453-4] Basic ideas and results of ray theory within the context of the theory of elasticity. Includes the necessary methods of applied mathematics, together with a complete discussion of the canonical problems. LCL

Applications (Physics), T(16-17: 1). Introduction to Mathematical Fluid Dynamics. Richard E. Meyer. Dover Pub, 1982, vi + 185 pp, $4.50 (P). [ISBN:0-486-61554-5] Reprint of 1971 text (TR, May 1972). Presupposes calculus and vector analysis; with problems and appendices included after most sections. RB

Applications (Physics), P. Percolation Theory for Mathematicians. Harry Kesten. Progress in Prob. & Stat., V. 2. Birkhauser Boston, 1982, 423 pp, $30. [ISBN: 3-7643-3107-0] A unified exposition of most of the known rigorous results concerning Bernoulli percolation in undirected graphs in two dimensions. Some new results are presented as well. AO

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Southern California Section

The Fall meeting of the Southern California Section was a joint meeting with SIAM held at Pepperdine University, Malibu, California on November 13, 1982. There were about 120 people in attendance.

Luncheon Talk:

"Information and the Uniqueness of Man," by Raymond Redheffer, University of California, Los Angeles.

Invited Addresses:

"Computer Memories, Hamming Codes, and the Birthday Surprise," by Robert Mc Eliece, Caltech.
"Applications of Functional Analysis and Complex Variables in Electrical Engineering," by William Helton, University of California, San Diego.
"Triptych of Singular Perturbation Problems," by Charles Lange, University of California, Los Angeles.
"Parameter Identifiability and Identification," by Stavros Busenberg, Harvey Mudd College.

Panel Discussion:

"Undergraduate Mathematics in Non-academic Employment." Roy Danchick, Abacus Programming (Moderator); Joe Wertz, Aerospace Corporation; Roger Bourke, Jet Propulsion Laboratory; Edward Posner, Caltech; Marvin Blum, Hughes Aircraft.

Louisiana-Mississippi Section

The Louisiana-Mississippi Section held its annual meeting at Delta State University, Cleveland, Mississippi on February 18-19, 1983. With the reduction in travel funds in both states, there were

100 registered faculty attendees and 17 student registrants. The meeting was the 60th joint meeting with the Louisiana-Mississippi branch of the NCTM.

Invited Addresses:

"Pressures for Curriculum Modification," by R.D. Anderson, Past President, MAA.
"Geometrical Proofs of Some Theorems About Differentiable Functions," by James R. Dorrah, Louisiana State University.

Student Papers:

"Force and Velocity in the English Frog Sartorius Muscle," by Bob Philpot, Belhaven College.
"Rose's Rose," by John A. Bailey, Millsaps College.
"Trapezoidial Properties," by Scott Bowie, Millsaps College.
"Sequential Verbalization for the Inverse of a One-to-One Function," by Laurel C. Eskridge, Millsaps College.
"Fibonacci Numbers and Patterns of Independent Sets in Cycles and Paths," by Mary Harrington, University of Mississippi.

Panel Discussions:

"Entrance Requirements for Public Universities," by A.J. Hulin, University of New Orleans and John L. Tilley, Mississippi State University.
"Mathematics Placement of Freshmen Students," by Nancy Dunn (Moderator), University of New Orleans.
"Junior College Transfer Students--Follow-Up," by Ken Johnston (Moderator), Hinds Junior College.

Contributed Papers:

"The Space of Entire Functions Revisited," by L. Swetharanyam, McNeese State University.
* "A Generalization of Euler's Phi-Function," by S. Leigh, University of Southwestern Louisiana.
* "Fixed Point Theorems for Certain Classes of Multi-Valued Mappings," by H. Kaneko, Mississippi State University.
"Fibonacci Numbers and Graphs," by W. Staton, University of Mississippi.
* "A Concrete Experiences Approach to Concept Formation in Developmental Arithmetic," by R. Yellott, McNeese State University.
"Microcomputers in the Junior High School Classroom," by R. Grantham, University of Mississippi.
"Microcomputer Aids in Teaching Trigonometry," by D. Cook, University of Mississippi.
* "Z-Closed Spaces," by T. Thompson, University of Southwestern Louisiana.
* "Semi-T_0-Identification Spaces and S-Essentially T_0 Spaces," by C. Dorsett, Louisiana Tech University.
* "Orderly Neighborhoods in a Disorderly Row," by M. Maxfield and J. Maxfield, Louisiana Tech University.
"A Discrimination Procedure for Bivariate Probability Models," by B. Asrabadi, Nicholls State University.

Student papers by John A. Bailey and Mary Harrington were awarded cash prizes of $50.

Southern California Section

The spring meeting of the Southern California Section was held on March 5, 1983 at the University of California, San Diego. There were approximately 200 people in attendance.

Invited Lectures:

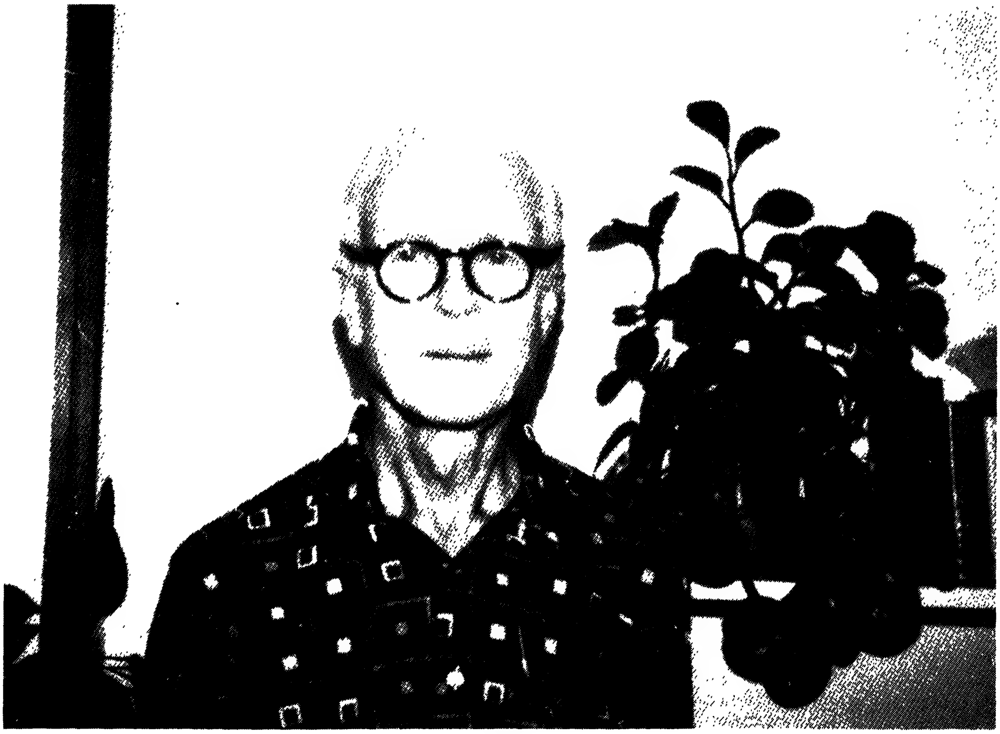
"Topology and Human Greed," by Michael Starbird, University of Texas, Austin.
"Coding Theory and Finite Geometrics," by Jacobus van Lint, Tech. University Eindhoven visiting Caltech.
"Can Probability be Applied to Mathematics," by Richard P. Feynmann, Caltech, Nobel Laureate.
"Synchronization for Deep Space Codes," by Laif Swanson, Jet Propulsion Laboratory, Caltech.

Report:

"Diagnostic Testing in Mathematics," by Philip P. Curtis, Jr., University of California, Los Angeles.

Luncheon Talk:

"Reminiscences in the History of Mathematics," by Ernst Straus, University of California, Los Angeles.



If any one man discovered differential topology, he did. See p. 397.

Lemma 1 we get

$$g - f = \sum_{n=1}^{\infty} (g_n - f_n).$$

Let $h_n = g - f - \sum_{m=1}^{n}(g_m - f_m)$; then $h_n \downarrow 0$ so $\Lambda(h_n) \to 0$, i.e.,

$$\nu([f, g)) = \sum_{n=1}^{\infty} \nu([f_n, g_n)).$$

As in the case of "ordinary" intervals, the proof that $\mathfrak{S}$ is a semiring is an easy exercise.

Proof of the theorem. For ν as defined in Lemma 3, choose ρ according to Lemma 2. Let $0 \leqslant f \in F$. For $f_n = (n[f - f \wedge 1]) \wedge 1$ we have

(A) $$[0, \beta f_n) \uparrow \{f > 1\} \times [0, \beta) \quad \text{for every } \beta > 0.$$

Hence,

$$\mu(A) = \rho(A \times [0,1)), A \in \mathfrak{B}_F$$

is a well-defined measure with the property

$$\rho(\{f > 1\} \times [0, \beta)) = \beta\mu(\{f > 1\}), f \in F, \beta > 0.$$

Let $B^n = \sum_{i=1}^{n2^n}(C_i^n - C_{i+1}^n) \times [0, i2^{-n})$ with $C_i^n = \{i^{-1}2^n f > 1\}$ for $1 \leqslant i \leqslant n2^n$ and $C_i^n = \varnothing$ for $i = n2^n + 1$. Then, $B^n \uparrow [0, f)$ and we get

$$\Lambda(f) = \rho([0, f)) = \lim_{n \to \infty} \rho(B^n) = \int f \, d\mu.$$

REMARKS.

(a) It is well known that one cannot dispense with "Stone's condition": $f \in F$ implies $f \wedge 1 \in F$. (Compare [**1**], 7XB for a counterexample.) However, any other condition which ensures for every $f \in F$ the existence of a sequence $f_n \in F$ with property (A) is also appropriate.

(b) Of course, the above theorem implies the Riesz representation theorem ([**3**], 56D). This is a well-known and immediate consequence of Dini's theorem. Another short proof for this special case has been given by Garling [**2**] (or [**4**], §22). Garling's method is closely related to Varadarajan's [**5**].

References

1. D. H. Fremlin, Topological Riesz Spaces and Measure Theory, Cambridge University Press, 1974.

2. D. J. H. Garling, A "short" proof of the Riesz representation theorem, Proc. Cambridge Philos. Soc., 73 (1973) 459–460.

3. P. R. Halmos, Measure Theory, Van Nostrand, Princeton, 1950.

4. R. B. Holmes, Geometric Functional Analysis and its Applications, Springer, New York-Heidelberg-Berlin, 1975.

5. V. S. Varadarajan, On a theorem of F. Riesz concerning the form of linear functionals, Fund. Math., XLVI (1958) 209–220.

ANSWER TO PHOTO ON PAGE 388

Hassler Whitney, one of the recipients of the 1982 Wolf Foundation Prize in Mathematics.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A CONJECTURE RELATED TO SYLVESTER'S PROBLEM

PETER BORWEIN AND MICHAEL EDELSTEIN

Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada

If a finite set of points spans the plane, then there exists a line through exactly two of the points. Posed as a problem by J. J. Sylvester in 1893, this was not solved until the 1930's when various people, including P. Erdős and T. Gallai, revived interest in it. Motzkin, one of a number of mathematicians to contribute to the now considerable list of refinements and variations on Sylvester's Problem, considered higher dimensional analogues. A substantial contribution to this higher dimensional problem was provided by Hansen in 1965. (See [5].) The rich literature on this problem may be traced through the references in [4], [6] and [7].

Of a distinctly similar flavor is another result of Motzkin. Given two finite and disjoint sets of points whose union spans the plane, then there exists a line through at least two points of one of the sets that does not intersect the other. Such a line is called monochrome. Essentially different proofs of this may be found in [4] and [9]. A beguilingly simple related question asks whether two finite and disjoint sets whose union spans R^4 must generate a monochrome plane (a plane through at least three noncollinear points of one of the sets that does not intersect the other set). This result is false in R^3 and may be easily proved in R^6 using the results in [5]. Another higher dimensional variation on Motzkin's theme is offered by Shannon [9]: if n finite and disjoint sets span projective n-space P^n, then there exists a line through at least two points of one of the sets that misses all the other sets. We propose the following conjecture that, if true, would provide an umbrella for the preceding results.

CONJECTURE. *If A and B are two finite and disjoint sets whose union spans P^{n+m}, then either there exists an A-monochrome n-flat or there exists a B-monochrome m-flat.*

By an A-monochrome n-flat we mean an affine variety of dimension n spanned by points of A that contains no points of B.

In [3] this conjecture is proved for $n = 1$. This, by induction, implies Shannon's result concerning n sets in n dimensions. There are various other immediate consequences that would follow from the conjecture. For instance, n sets in $2n$ dimensions would guarantee the existence of a monochrome plane.

In [2] it is shown that any finite set E that spans $P^{2(n+m)}$ generates an $(n + m)$-flat spanned by and containing exactly $n + m + 1$ points of E. Thus, the conclusion of the above conjecture is true if we replace P^{n+m} by $P^{2(n+m)}$. (As is often the case in this area, finding some dimension in which the problem is tractable is much simpler than determining the "correct" dimension.) If we consider $n + m$ points in general position in P^{n+m-1} and let A be composed of n of these points and let B be composed of the remaining m points, we see that we cannot hope to replace P^{n+m} by P^{n+m-1} without violating the conjecture.

An apparently related question concerns compact sets in R^{n+m} which are also countable. More precisely, if we assume that A and B are countable and compact rather than just finite, does the

conjecture still hold? If not, what dimension, if any, suffices to maintain the conclusion?

As far as we know, this question is only resolved in the case of two sets in the plane. However, Tingley [10] shows that two compact disjoint sets in R^3 generate a monochrome line. Perhaps by stepping up a dimension the condition that the sets be countable can in general be dropped.

A curious feature of a number of results related to the conjecture is that they seem to be more easily proved in dual formulation. The latter questions concerning compact sets do not dualize as naturally. While the easiest arguments for treating Motzkin and Sylvester-like problems for finite sets are geometrical, the proofs of results for infinite sets are topological in nature. A perhaps surprising feature of many of these problems is that combinatorial methods do not prove particularly useful.

References

1. V. J. Baston and F. A. Bostock, A Gallai-type problem, J. Combin. Theory, Ser. A, 24 (1978) 122–125.
2. W. E. Bonnice and M. Edelstein, Flats associated with finite sets in P^d, Nieuw Arch. Wisk., 15 (1967) 11–14.
3. P. Borwein, On monochrome lines and hyperplanes, J. Combin. Theory, Ser. A, 33 (1982) 76–81.
4. G. D. Chakerian, Sylvester's problem on collinear points and a relative, this MONTHLY, 77 (1970) 164–167.
5. S. Hansen, A generalization of a theorem of Sylvester on the lines determined by a finite point set, Math. Scand, 16 (1965) 175–180.
6. S. Hansen, Contributions to the Sylvester-Gallai theory, Ph.D. thesis, Kobenhavns Universitet, 1981.
7. W. Moser, Research problems in discrete geometry, McGill Mathematics Report 81-3.
8. T. S. Motzkin, Nonmixed connecting lines, Notices Amer. Math. Soc., 14 (1967) 837.
9. R. Shannon, Ph.D. Thesis, University of Washington (1974).
10. D. Tingley, M. A. Thesis, Dalhousie University (1976).

MISCELLANEA

106. *Metagame* (so baptized by Martin Gardner) is an ingenious new twist on the Russell paradox. (It was called "hypergame" by its discoverer, William S. Zwicker.) Here is how it goes.

A two-person game (such as tic-tac-toe or chess) is *finite* if it always ends in a finite number of moves. The first move of metagame is to pick a finite game. If, for example, you and I are playing metagame and I have the first move, I can say "Let's play chess." Then you make the first move in a chess game, and we continue playing till that game ends.

Is metagame finite? If it is, then, as my first move in a metagame, I can say "Let's play metagame." It is now your turn, and, as your first move in the metagame that I chose to play, you can say "Let's play metagame." The process can obviously go on indefinitely, contrary to the assumption that metagame is finite. The contradiction forces the conclusion that metagame is not finite. It follows that I cannot choose metagame as my first move in metagame; I must choose a finite game. Consequence: the game must end in a finite number of moves, and that contradicts the proved statement that metagame is not finite.

—Paraphrased, by permission, from a
forthcoming book by Raymond Smullyan.

conjecture still hold? If not, what dimension, if any, suffices to maintain the conclusion?

As far as we know, this question is only resolved in the case of two sets in the plane. However, Tingley [10] shows that two compact disjoint sets in R^3 generate a monochrome line. Perhaps by stepping up a dimension the condition that the sets be countable can in general be dropped.

A curious feature of a number of results related to the conjecture is that they seem to be more easily proved in dual formulation. The latter questions concerning compact sets do not dualize as naturally. While the easiest arguments for treating Motzkin and Sylvester-like problems for finite sets are geometrical, the proofs of results for infinite sets are topological in nature. A perhaps surprising feature of many of these problems is that combinatorial methods do not prove particularly useful.

References

1. V. J. Baston and F. A. Bostock, A Gallai-type problem, J. Combin. Theory, Ser. A, 24 (1978) 122–125.
2. W. E. Bonnice and M. Edelstein, Flats associated with finite sets in P^d, Nieuw Arch. Wisk., 15 (1967) 11–14.
3. P. Borwein, On monochrome lines and hyperplanes, J. Combin. Theory, Ser. A, 33 (1982) 76–81.
4. G. D. Chakerian, Sylvester's problem on collinear points and a relative, this MONTHLY, 77 (1970) 164–167.
5. S. Hansen, A generalization of a theorem of Sylvester on the lines determined by a finite point set, Math. Scand, 16 (1965) 175–180.
6. S. Hansen, Contributions to the Sylvester-Gallai theory, Ph.D. thesis, Kobenhavns Universitet, 1981.
7. W. Moser, Research problems in discrete geometry, McGill Mathematics Report 81-3.
8. T. S. Motzkin, Nonmixed connecting lines, Notices Amer. Math. Soc., 14 (1967) 837.
9. R. Shannon, Ph.D. Thesis, University of Washington (1974).
10. D. Tingley, M. A. Thesis, Dalhousie University (1976).

MISCELLANEA

106. *Metagame* (so baptized by Martin Gardner) is an ingenious new twist on the Russell paradox. (It was called "hypergame" by its discoverer, William S. Zwicker.) Here is how it goes.

A two-person game (such as tic-tac-toe or chess) is *finite* if it always ends in a finite number of moves. The first move of metagame is to pick a finite game. If, for example, you and I are playing metagame and I have the first move, I can say "Let's play chess." Then you make the first move in a chess game, and we continue playing till that game ends.

Is metagame finite? If it is, then, as my first move in a metagame, I can say "Let's play metagame." It is now your turn, and, as your first move in the metagame that I chose to play, you can say "Let's play metagame." The process can obviously go on indefinitely, contrary to the assumption that metagame is finite. The contradiction forces the conclusion that metagame is not finite. It follows that I cannot choose metagame as my first move in metagame; I must choose a finite game. Consequence: the game must end in a finite number of moves, and that contradicts the proved statement that metagame is not finite.

—Paraphrased, by permission, from a
forthcoming book by Raymond Smullyan.

② NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, NM 55057.

POINTWISE LIMITS OF ANALYTIC FUNCTIONS

KENNETH R. DAVIDSON
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada

What are the possible pointwise limits of analytic functions? Since analytic functions are well behaved, you might expect such limits to be "nice" also. Yet it turns out that some bizarre things can occur. The results in this paper are accessible to a student of complex analysis, but they do not appear to have been collected together before.

Classical Background. It is easy to construct a sequence of continuous functions which converges pointwise to a function with discontinuities. However, not every function is the pointwise limit of continuous functions. The limit of a sequence of continuous functions which converges uniformly is necessarily continuous. We will be considering continuous functions on an open subset Ω of the complex plane. In this context, it is more natural to consider a topology of convergence in between uniform and pointwise convergence—*uniform convergence on compact subsets* (u.c.c.). A sequence of functions f_n on Ω converges u.c.c. if for every compact subset K of Ω, the restrictions $f_n|K$ converge uniformly. If every f_n is continuous, then the limit function is continuous on each K and thus is continuous on Ω. If every f_n is analytic, the u.c.c. limit is analytic also.

In both real and complex analysis, mathematicians have looked for and found conditions on a family of functions $\mathfrak{F}$ on Ω which guarantee that sequences taken from $\mathfrak{F}$ have convergent subsequences. These results are closely related to our problem. To explain the classical theorems, we need some additional terminology.

Recall that a function is continuous at z_0 if for each $\varepsilon > 0$, there is a $\delta > 0$ so that $|z - z_0| < \delta$ implies $|f(z) - f(z_0)| < \varepsilon$. A family of functions $\mathfrak{F}$ on Ω is *equicontinuous* if for each z_0 in Ω and each $\varepsilon > 0$ the $\delta > 0$ given above can be chosen independent of f in $\mathfrak{F}$. That is, $|z - z_0| < \delta$ implies $|f(z) - f(z_0)| < \varepsilon$ for all f in $\mathfrak{F}$.

A family of functions $\mathfrak{F}$ is *pointwise bounded* if for each z_0 in Ω, the set $\{f(z_0): f \in \mathfrak{F}\}$ is bounded. The family $\mathfrak{F}$ is *locally bounded* if each z_0 in Ω is contained in an open ball U so that the set $\{f(z): z \in U, f \in \mathfrak{F}\}$ is bounded. Since a convergent sequence is bounded, a sequence of continuous functions converging pointwise is pointwise bounded. However, it need not satisfy the stronger condition of local boundedness. A sequence converging u.c.c., though, must be locally bounded because every z_0 is contained in a ball U on which the sequence converges uniformly.

The following classical theorem from real analysis is proved in many introductory texts.

THEOREM A (*Arzela-Ascoli*). *Let $\mathfrak{F}$ be a family of continuous functions on Ω. Every sequence in $\mathfrak{F}$ has a subsequence converging u.c.c. if and only if $\mathfrak{F}$ is equicontinuous and pointwise bounded.*

The analogous theorem for analytic functions is:

THEOREM M (*Montel*). *Let $\mathfrak{F}$ be a family of analytic functions on Ω. Every sequence in $\mathfrak{F}$ has a subsequence converging u.c.c. if and only if $\mathfrak{F}$ is locally bounded.*

Notice that the equicontinuity condition is eliminated at the price of a stronger boundedness condition. A natural proof of Montel's Theorem is obtained by deducing equicontinuity and applying Theorem A.

Sketch. Fix w_0 in Ω and let U be a ball of radius r about w_0 so that $|f(w)| \leq M$ for all w in U and f in $\mathcal{F}$. Let $\mathcal{C}$ be the circle of radius $r/2$ centered at w_0. For every w within $r/4$ of w_0, Cauchy's Theorem gives

$$f(w) - f(w_0) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f(z)}{z-w} - \frac{f(z)}{z-w_0}\,dz = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f(z)\,dz}{(z-w)(z-w_0)}(w-w_0).$$

A simple estimate gives $|f(w) - f(w_0)| \leq 4Mr^{-1}|w - w_0|$ for all f in $\mathcal{F}$. From this, equicontinuity is immediate. □

What happens when a family of analytic functions is only pointwise bounded? As we noted earlier, this question applies in particular to a sequence of analytic functions which converges pointwise on Ω. The answers are dramatically different than for Montel's Theorem. There are some positive results, but first we will examine some of the pathology. Another classical result, Runge's Theorem, is needed to provide a way of obtaining analytic functions with prescribed behavior.

THEOREM R (*Runge*). *Let K be a compact subset of the complex plane with connected complement. Let f be a function analytic in a neighborhood of K. Then there is a sequence of polynomials which converges to f uniformly on K.*

Sketch. Inside the neighborhood U of K on which f is analytic, it is possible to draw a finite collection $\mathcal{C}$ of piecewise smooth curves which surround K in such a way that Cauchy's Theorem is valid for f. That is,

$$f(z) = \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{f(w)}{w-z}\,dw \quad \text{for all } z \text{ in } K.$$

This integral can be approximated by Riemann sums $\Sigma_{j=1}^{n} f(w_j)(w_{j+1} - w_j)/(w_j - z)$. Since $\mathcal{C}$ is bounded away from K, simple estimates show that these sums converge uniformly on K. This approximates f by rational functions. In order to get polynomials, it suffices to approximate each $g(z) = (w - z)^{-1}$ by polynomials uniformly on K.

When $|w|$ is strictly larger than $|z|$ for all z in K, the geometric series $\Sigma_{n=0}^{\infty} z^n/w^{n+1}$ converges uniformly to $(w - z)^{-1}$ on K. A similar argument can be used to express $(w' - z)^{-1}$ in terms of powers of $(w - z)^{-1}$ for w' near w. Thus if $(w - z)^{-1}$ can be approximated by polynomials in K, so can $(w' - z)^{-1}$. Because $\mathbb{C} \setminus K$ is a connected open set, a finite number of repetitions will get us from the "big" values of w to any point in $\mathbb{C} \setminus K$. Thus a polynomial approximation is obtained. □

Some Examples.

EXAMPLE 1. Let K_n be the union of the point $\{0\}$, the line segment $[(1/n), n]$ and the compact set $S_n = \{z \in \mathbb{C}: |z| \leq n \text{ and } \operatorname{dist}(z, \mathbb{R}_+) \geq 1/n\}$. (See Diagram 1 on p. 393.) Let g_n be an analytic function which vanishes in a neighborhood of S_n and $[(1/n), n]$ and is constantly one in a ball about $\{0\}$. Let p_n be a polynomial obtained by using Runge's Theorem such that $|p_n(z) - g_n(z)| < 1/n$ for all z in K_n.

Every point z in the complex plane belongs to K_n for large enough n. Thus $\lim_{n\to\infty} p_n(z)$ exists pointwise and equals zero everywhere except at $\{0\}$ where the limit is one! This limit is uniform on each set S_n by construction. Hence the convergence is u.c.c. on $\mathbb{C} \setminus [0, \infty]$. The sequence $\{p_n\}$ cannot converge uniformly on any neighborhood of $\{0\}$ because the limit function is not continuous at $\{0\}$. In fact, it cannot converge uniformly near any point on the positive real axis.

To see this, notice that if $\{p_n\}$ converges uniformly on a ball U of radius r centered at some $x > 0$, then for some $n \geq 2$, $r > 1/n$ and $|p_n(z)| < 1/2$ on U. In addition, $|p_n(z)| < 1/n \leq 1/2$ on S_n. But the union of S_n and U contains the circle of radius x about zero. So the maximum modulus principle implies that $|p_n(0)| < 1/2$ which is a contradiction. □

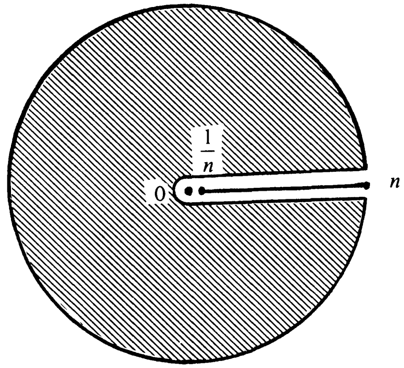


DIAGRAM 1

EXAMPLE 2. Let L_n be the union of K_n and the point $\{1/2n\}$. Let h_n be an analytic function on a neighborhood of L_n which is zero near K_n and one at $\{1/2n\}$. Again, Runge's Theorem provides polynomials q_n such that $|q_n(z) - h_n(z)| < 1/n$ on L_n. Clearly, the sequence $\{q_n\}$ converges pointwise to zero everywhere. However, convergence is not uniform near $\{0\}$ because $q_n(1/2n)$ is approximately equal to one. As in Example 1, convergence is uniform on each S_n but cannot converge uniformly near any point on the positive real axis. □

Some Positive Results.

PROPOSITION 1. *Let $\mathcal{F}$ be a pointwise bounded family of analytic functions on Ω. There is a (maximal) dense open set $\Omega_0 \subseteq \Omega$ such that $\mathcal{F}$ restricted to Ω_0 is locally bounded.*

Proof. For each z in Ω, let $\phi(z)$ denote the least upper bound for $\{|f(z)|: f \in \mathcal{F}\}$. Because each f in $\mathcal{F}$ is continuous, it is easy to see that $K_n = \{z: \phi(z) \leq n\}$ is (relatively) closed in Ω. Since $\phi(z)$ is finite for each z, Ω is the union of the K_n. The Baire Category Theorem shows that for every ball U in Ω, $K_n \cap U$ has interior for large n. Thus, the union Ω_0 of the interiors int K_n is a dense open subset of Ω. Obviously, $\mathcal{F}$ is bounded on each int K_n, and hence $\mathcal{F}$ is locally bounded on Ω_0. Conversely, if $\mathcal{F}$ is bounded by n on an open set U, then U is contained in int K_n and thus is in Ω_0. □

COROLLARY 2. *Let f_n be a sequence of analytic functions converging pointwise on Ω. Then the limit function f is analytic on a dense open subset of Ω.*

Proof. A sequence of functions converging pointwise is pointwise bounded. By Proposition 1, $\{f_n\}$ is locally bounded on Ω_0. By Montel's Theorem, a subsequence converges u.c.c. to f on Ω_0 and hence f is analytic on Ω_0. □

PROPOSITION 3. *If f_n is a sequence of bounded analytic functions on Ω converging pointwise to f, then f is analytic and the convergence is u.c.c..*

Proof. The family $\mathcal{F} = \{f_n\}$ is bounded and thus locally bounded. By Montel's Theorem, every subsequence of f_n has a sub-subsequence which converges u.c.c. This limit is f perforce, so f is analytic. If the whole sequence does not converge to f u.c.c., there would be a compact set K in Ω, an $\varepsilon > 0$, a subsequence f_{n_k}, and points z_k in K such that $|f_{n_k}(z_k) - f(z_k)| \geq \varepsilon$ for $k \geq 1$. Thus no subsequence of f_{n_k} could converge uniformly to f on K, which is a contradiction. □

What Limits Are Possible? How can the set of pointwise limits of analytic functions be described? If f is such a limit, then by Proposition 1 there is a dense open set Ω_0 on which the convergence is u.c.c. and f is analytic. On the relatively closed, nowhere dense set $C = \Omega \setminus \Omega_0$, f is merely the pointwise limit of continuous functions. That is all that can be said, for a modification of Example 1 can be arranged to converge to any function with these properties.

The idea is to choose increasing sequences K_n and C_n of subsets of Ω_0 and C respectively so that the complement of $K_n \cup C_n$ is connected and the union of all K_n's and C_n's is Ω.

Use Runge's theorem to approximate f within $1/n$ on K_n by polynomials. On C, there is a sequence g_n of continuous functions converging pointwise to f. A theorem of Lavrentiev can be used to approximate g_n within $1/n$ on C_n by a polynomial. Finally, use Runge's Theorem again to approximate both these polynomials simultaneously by a polynomial f_n. It is clear that $\{f_n\}$ converges to f.

A careful choice of the sequence g_n, as in Example 2, allows one to ensure that the convergence is not uniform near any point of C. Thus one may achieve an analytic limit f with failure to converge u.c.c. on any given closed nowhere dense subset.

Sequences of univalent functions (one to one and analytic) are much more nicely behaved. A sequence of univalent functions converging pointwise to a nonconstant function must in fact converge u.c.c.. In particular, the limit is always analytic. These facts will not be proved here. We just mention that the proof relies on a deep theorem of Montel which states that a family of analytic functions which omits two values is locally bounded.

References

The Arzela-Ascoli Theorem can be found in *Elementary Classical Analysis* by Marsden, and in greater generality in *General Topology* by Kelley. Montel's Theorem and Runge's Theorem can be found in most texts on complex analysis, for example, *Functions of One Complex Variables* by Conway. Another proof of Proposition 3 is contained in *A Second Course in Complex Variables* by Veech. Examples related to ours are contained in the problems section of *Real and Complex Analysis*, 2nd ed., by Rudin (pages 294 and 359). Lavrentiev's Theorem is a special case of Mergelyan's Theorem, which can be found in the last chapter of Rudin's book. The theorem of Montel referred to in the last paragraph is in Chapter 4 of Veech.

A NET OF EXPONENTIALS CONVERGING TO A NONMEASURABLE FUNCTION

LEE A. RUBEL AND ARISTOMENIS SISKAKIS
Department of Mathematics, University of Illinois, Urbana IL 61801

It is a basic fact that the pointwise limit of a *sequence* of (Lebesgue) measurable functions must be measurable. It is not hard to see that when "sequence" is replaced by "net," the result fails. Nevertheless, it is surprising that from the family $T = \{f(x - t) : t \in \mathbb{R}\}$ of translates of $f(x) = \exp(ix^2)$, a pointwise convergent net may be drawn whose limit is nonmeasurable. This is just another way of saying that in the space of all complex-valued functions on $\mathbb{R}$, in the topology of pointwise convergence, the closure of T contains a nonmeasurable function. (For an extended discussion of nets and filters in topology, see [**BAR**]. Essentially, the same fact that we prove here is proved in [**LUS**, Example 8.4.45, p. 218] by other means.)

To prove that there exists such a net of translates of $\exp(ix^2)$, let n_k be a sequence of positive integers approaching ∞. There are several ways to see that there exists a subnet (n_{k_γ}) of (n_k) such that for any bounded continuous function g on $\mathbb{R}$, $\phi(g) = \lim_\gamma g(n_{k_\gamma})$ exists. In one kind of language, ϕ is a point of the Stone-Čech compactification of $\mathbb{R}$—see [**GIJ**, Chapter 6] for more details. From another viewpoint, ϕ arises as a weak-star limit point of the functionals of evaluation (on the Banach space of continuous bounded functions on $\mathbb{R}$) at the points n_k. In any case, it is clear that ϕ is a complex homomorphism—that is, ϕ takes complex values and $\phi(g + h) = \phi(g) + \phi(h)$ and $\phi(gh) = \phi(g) \cdot \phi(h)$. To stress the dependence on x as the running variable, we shall write $\phi_x(f(x))$ sometimes. We will see that if the sequence (n_k) satisfies property P below, then $\phi_x(f(x + t)) = \lim_\gamma f(t + n_{k_\gamma})$ is not a measurable function of t.

With $f(x) = \exp(ix^2)$ we have

$$\phi_x(f(x + t)) = \phi_x(e^{i(x+t)^2}) = \phi_x(e^{ix^2})\phi_x(e^{it^2})\phi_x(e^{2ixt}).$$

The idea is to choose increasing sequences K_n and C_n of subsets of Ω_0 and C respectively so that the complement of $K_n \cup C_n$ is connected and the union of all K_n's and C_n's is Ω.

Use Runge's theorem to approximate f within $1/n$ on K_n by polynomials. On C, there is a sequence g_n of continuous functions converging pointwise to f. A theorem of Lavrentiev can be used to approximate g_n within $1/n$ on C_n by a polynomial. Finally, use Runge's Theorem again to approximate both these polynomials simultaneously by a polynomial f_n. It is clear that $\{f_n\}$ converges to f.

A careful choice of the sequence g_n, as in Example 2, allows one to ensure that the convergence is not uniform near any point of C. Thus one may achieve an analytic limit f with failure to converge u.c.c. on any given closed nowhere dense subset.

Sequences of univalent functions (one to one and analytic) are much more nicely behaved. A sequence of univalent functions converging pointwise to a nonconstant function must in fact converge u.c.c.. In particular, the limit is always analytic. These facts will not be proved here. We just mention that the proof relies on a deep theorem of Montel which states that a family of analytic functions which omits two values is locally bounded.

References

The Arzela-Ascoli Theorem can be found in *Elementary Classical Analysis* by Marsden, and in greater generality in *General Topology* by Kelley. Montel's Theorem and Runge's Theorem can be found in most texts on complex analysis, for example, *Functions of One Complex Variables* by Conway. Another proof of Proposition 3 is contained in *A Second Course in Complex Variables* by Veech. Examples related to ours are contained in the problems section of *Real and Complex Analysis*, 2nd ed., by Rudin (pages 294 and 359). Lavrentiev's Theorem is a special case of Mergelyan's Theorem, which can be found in the last chapter of Rudin's book. The theorem of Montel referred to in the last paragraph is in Chapter 4 of Veech.

A NET OF EXPONENTIALS CONVERGING TO A NONMEASURABLE FUNCTION

LEE A. RUBEL AND ARISTOMENIS SISKAKIS
Department of Mathematics, University of Illinois, Urbana IL 61801

It is a basic fact that the pointwise limit of a *sequence* of (Lebesgue) measurable functions must be measurable. It is not hard to see that when "sequence" is replaced by "net," the result fails. Nevertheless, it is surprising that from the family $T = \{f(x - t) : t \in \mathbb{R}\}$ of translates of $f(x) = \exp(ix^2)$, a pointwise convergent net may be drawn whose limit is nonmeasurable. This is just another way of saying that in the space of all complex-valued functions on $\mathbb{R}$, in the topology of pointwise convergence, the closure of T contains a nonmeasurable function. (For an extended discussion of nets and filters in topology, see [**BAR**]. Essentially, the same fact that we prove here is proved in [**LUS**, Example 8.4.45, p. 218] by other means.)

To prove that there exists such a net of translates of $\exp(ix^2)$, let n_k be a sequence of positive integers approaching ∞. There are several ways to see that there exists a subnet (n_{k_γ}) of (n_k) such that for any bounded continuous function g on $\mathbb{R}$, $\phi(g) = \lim_\gamma g(n_{k_\gamma})$ exists. In one kind of language, ϕ is a point of the Stone-Čech compactification of $\mathbb{R}$—see [**GIJ**, Chapter 6] for more details. From another viewpoint, ϕ arises as a weak-star limit point of the functionals of evaluation (on the Banach space of continuous bounded functions on $\mathbb{R}$) at the points n_k. In any case, it is clear that ϕ is a complex homomorphism—that is, ϕ takes complex values and $\phi(g + h) = \phi(g) + \phi(h)$ and $\phi(gh) = \phi(g) \cdot \phi(h)$. To stress the dependence on x as the running variable, we shall write $\phi_x(f(x))$ sometimes. We will see that if the sequence (n_k) satisfies property P below, then $\phi_x(f(x + t)) = \lim_\gamma f(t + n_{k_\gamma})$ is not a measurable function of t.

With $f(x) = \exp(ix^2)$ we have

$$\phi_x(f(x + t)) = \phi_x(e^{i(x+t)^2}) = \phi_x(e^{ix^2})\phi_x(e^{it^2})\phi_x(e^{2ixt}).$$

Now $\phi_x(\exp(ix^2))$ is just a complex number. It is nonzero since $\exp(in_{k_\gamma}^2)$ has modulus 1 for all γ. Since $\exp(it^2)$ is a constant function with respect to x, we have $\phi_x(\exp(it^2)) = \exp(it^2)$, which is a measurable function of t. So we can prove that $\phi_x(f(x+t))$ is nonmeasurable if we can prove that $\nu(t) = \phi_x(\exp(2ixt))$ is nonmeasurable.

It is immediate that ν satisfies the functional equation

$$\nu(s+t) = \nu(s)\nu(t) \tag{1}$$

and that $|\nu(t)| = 1$ for all t, and $\nu(t)$ is π-periodic so we can use the well-known fact that unless

$$\nu(t) = e^{i\lambda t} \tag{2}$$

for some constant λ, then ν must be nonmeasurable. (A folklore proof of this is the following: Suppose ν were measurable. The convolution

$$(u * v)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(x-t)v(t)\,dt$$

(with the usual abuse of circular notation) of two bounded measurable functions is continuous, but

$$(\nu * \nu)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\nu(x)}{\nu(t)}\cdot\nu(t)\,dt = \nu(x)$$

so that ν itself would be continuous. Then (2) follows in a few easy lines by studying first rational values of s and t in (1) and then using the continuity of ν.)

PROPERTY P. *There exist two numbers t_0 and t_1 in $\mathbb{R}$ that are linearly independent over the rational numbers $\mathbb{Q}$, and two sequences (l_k) and (m_k) of positive integers, and sequences (ε_k) and (δ_k) of real numbers with $\varepsilon_k \to 0$ and $\delta_k \to 0$ as $k \to \infty$, so that*

$$\frac{t_0}{\pi} - \frac{l_k}{n_k} = \frac{\delta_k}{2n_k} \tag{3}$$

$$\frac{t_1}{\pi} - \frac{2m_k+1}{2n_k} = \frac{\varepsilon_k}{2n_k}. \tag{4}$$

Let us suppose for a minute that we have constructed our sequence (n_k) to have property P. Then the rest is easy. We have (with the assumption that $\nu(t)$ is measurable)

$$e^{i2n_{k_\gamma}t} \to e^{i\lambda t} \quad \text{for all } t.$$

Taking $t = t_0$ and then $t = t_1$ and solving (3) and (4) for n_k gives

$$e^{i\pi(2l_{k_\gamma}+\delta_{k_\gamma})} \to e^{i\lambda t_0}$$
$$e^{i\pi((2m_{k_\gamma}+1)+\varepsilon_{k_\gamma})} \to e^{i\lambda t_1}.$$

Hence $\exp(i\lambda t_0) = +1$ and $\exp(i\lambda t_1) = -1$, so that $\lambda t_0 = 2m\pi$, $\lambda t_1 = (2l+1)\pi$ for some integers l and m, and thus $t_0/t_1 \in \mathbb{Q}$, a contradiction.

To see that the property P can hold, let $n_k \to \infty$ very rapidly through even values, say $n_k = (k!)!$, $k = 2, 3, \ldots$. Let

$$\frac{t_0}{\pi} = \sum_{k=2}^{\infty}\frac{1}{n_k}, \quad \frac{t_1}{\pi} = \sum_{k=2}^{\infty}\frac{(-1)^{\alpha_k}}{2n_k},$$

where, for each k, $\alpha_k = 0$ or $\alpha_k = 1$. Let

$$\frac{l_k}{n_k} = \sum_{j=2}^{k}\frac{1}{n_j}, \quad \frac{2m_k+1}{2n_k} = \sum_{j=2}^{k}\frac{(-1)^{\alpha_j}}{2n_j},$$

and let δ_k and ε_k be defined by (3) and (4). It is easy to check that these formulas define integers l_k

and m_k and that $\delta_k \to 0$ and $\varepsilon_k \to 0$. Finally, there are uncountably many choices of the sequence (α_k), each one of them leading to a different t_1. Since there are only countably many t_1 that are rationally dependent on t_0, the result is proved.

As a final remark, we tried, but were unable to prove from our result that there is a net of translates of $\sin(x^2)$ that converges to a nonmeasurable function. Could there actually be a net $(\sin(x - t_\gamma)^2)$ that converges to a measurable function while $(\cos(x - t_\gamma)^2)$ converges to a nonmeasurable function?

References

[BAR] R. G. Bartle, Nets and filters in topology, this MONTHLY, 62 (1955) 551–557 (Correction in 70 (1963) 52–53).

[GIJ] L. Gillman and M. Jerison, Rings of Continuous Functions, Van Nostrand, Princeton, 1960.

[LUS] W. A. J. Luxemburg and K. Stroyan, Introduction to the Theory of Infinitesimals, Academic Press, New York, 1976.

A SIMPLE PROOF OF THE DANIELL-STONE REPRESENTATION THEOREM

JÜRGEN KINDLER
Department of Mathematics, TH Darmstadt, Schlossgartenstr. 7, 6100 Darmstadt, West Germany

Let X be a nonvoid set, F a linear space of real valued functions on X such that $f, g \in F$ implies $f \wedge g \in F$ and $f \wedge 1 \in F$ ($(f \wedge g)(x) = \min\{f(x), g(x)\}$); let $\mathfrak{B}_F$ denote the smallest σ-ring which contains the sets $\{f > 1\} = \{x \in X\colon f(x) > 1\}, f \in F$; and let $\Lambda\colon F \to \mathbb{R}$ be a positive linear functional such that $f_n \downarrow 0$ (i.e., $f_1(x) \geqslant f_2(x) \geqslant \cdots$ and $\lim_{n\to\infty} f_n(x) = 0$ for every $x \in X$) implies $\lim_{n\to\infty}\Lambda(f_n) = 0$.

We present a short and simple proof of the following fundamental theorem:

THEOREM (*Daniell-Stone*). *There exists a measure μ on $\mathfrak{B}_F$ such that $\Lambda(f) = \int f d\mu$ for every $f \in F$.*

In our proof we follow the same lines which are known from the construction of Lebesgue measure. Both constructions are based on the following two well-known facts (compare [3], §8D, §8 (5), and §13A). The union of disjoint sets $A_1, A_2, \cdots$ is denoted by $\sum_{n=1}^{\infty} A_n$.

LEMMA 1. *For all sequences $a_n, b_n \in \mathbb{R}$ with $b_n \geqslant a_n$, $n \geqslant 0$ and $[a_0, b_0) = \sum_{n=1}^{\infty}[a_n, b_n)$ we have $b_0 - a_0 = \sum_{n=1}^{\infty}(b_n - a_n)$.*

LEMMA 2. *Every σ-additive nonnegative finite set function ν on a semiring $\mathfrak{S}$ can be extended to a measure ρ on the σ-ring generated by $\mathfrak{S}$.*

Let us recall that a semiring is a nonvoid system $\mathfrak{S}$ of subsets of a fixed set with the properties

(i) if $S \in \mathfrak{S}$ and $T \in \mathfrak{S}$, then $S \cap T \in \mathfrak{S}$, and
(ii) if $S \in \mathfrak{S}$ and $T \in \mathfrak{S}$ and $S \subset T$, then there is a finite class $\{S_0, S_1, \ldots, S_n\}$ of sets in $\mathfrak{S}$ such that $S = S_0 \subset S_1 \subset \cdots \subset S_n = T$ and $R_i = S_i - S_{i-1} \in \mathfrak{S}$ for $i = 1, \ldots, n$.

Lemma 2 will be applied to the following situation:

LEMMA 3. *The system $\mathfrak{S}$ of "intervals"*

$$[f, g) = \{(x, t) \in X \times \mathbb{R} : f(x) \leqslant t < g(x)\}, \quad g \geqslant f, \quad g, f \in F$$

is a semiring and

$$\nu\colon \mathfrak{S} \to \mathbb{R}_+, \quad \nu([f, g)) = \Lambda(g - f)$$

is well-defined and σ-additive.

Proof. $[f, g) = \sum_{n=1}^{\infty}[f_n, g_n)$ implies $[f(x), g(x)) = \sum_{n=1}^{\infty}[f_n(x), g_n(x))$ for every $x \in X$. From

and m_k and that $\delta_k \to 0$ and $\varepsilon_k \to 0$. Finally, there are uncountably many choices of the sequence (α_k), each one of them leading to a different t_1. Since there are only countably many t_1 that are rationally dependent on t_0, the result is proved.

As a final remark, we tried, but were unable to prove from our result that there is a net of translates of $\sin(x^2)$ that converges to a nonmeasurable function. Could there actually be a net $(\sin(x - t_\gamma)^2)$ that converges to a measurable function while $(\cos(x - t_\gamma)^2)$ converges to a nonmeasurable function?

References

[BAR] R. G. Bartle, Nets and filters in topology, this MONTHLY, 62 (1955) 551–557 (Correction in 70 (1963) 52–53).

[GIJ] L. Gillman and M. Jerison, Rings of Continuous Functions, Van Nostrand, Princeton, 1960.

[LUS] W. A. J. Luxemburg and K. Stroyan, Introduction to the Theory of Infinitesimals, Academic Press, New York, 1976.

A SIMPLE PROOF OF THE DANIELL-STONE REPRESENTATION THEOREM

JÜRGEN KINDLER

Department of Mathematics, TH Darmstadt, Schlossgartenstr. 7, 6100 Darmstadt, West Germany

Let X be a nonvoid set, F a linear space of real valued functions on X such that $f, g \in F$ implies $f \wedge g \in F$ and $f \wedge 1 \in F$ ($(f \wedge g)(x) = \min\{f(x), g(x)\}$); let $\mathfrak{B}_F$ denote the smallest σ-ring which contains the sets $\{f > 1\} = \{x \in X\colon f(x) > 1\}, f \in F$; and let $\Lambda\colon F \to \mathbb{R}$ be a positive linear functional such that $f_n \downarrow 0$ (i.e., $f_1(x) \geqslant f_2(x) \geqslant \cdots$ and $\lim_{n\to\infty} f_n(x) = 0$ for every $x \in X$) implies $\lim_{n\to\infty}\Lambda(f_n) = 0$.

We present a short and simple proof of the following fundamental theorem:

THEOREM (*Daniell-Stone*). *There exists a measure μ on $\mathfrak{B}_F$ such that $\Lambda(f) = \int f d\mu$ for every $f \in F$.*

In our proof we follow the same lines which are known from the construction of Lebesgue measure. Both constructions are based on the following two well-known facts (compare [3], §8D, §8 (5), and §13A). The union of disjoint sets $A_1, A_2, \cdots$ is denoted by $\Sigma_{n=1}^{\infty} A_n$.

LEMMA 1. *For all sequences $a_n, b_n \in \mathbb{R}$ with $b_n \geqslant a_n$, $n \geqslant 0$ and $[a_0, b_0) = \Sigma_{n=1}^{\infty}[a_n, b_n)$ we have $b_0 - a_0 = \Sigma_{n=1}^{\infty}(b_n - a_n)$.*

LEMMA 2. *Every σ-additive nonnegative finite set function ν on a semiring $\mathfrak{S}$ can be extended to a measure ρ on the σ-ring generated by $\mathfrak{S}$.*

Let us recall that a semiring is a nonvoid system $\mathfrak{S}$ of subsets of a fixed set with the properties

(i) if $S \in \mathfrak{S}$ and $T \in \mathfrak{S}$, then $S \cap T \in \mathfrak{S}$, and
(ii) if $S \in \mathfrak{S}$ and $T \in \mathfrak{S}$ and $S \subset T$, then there is a finite class $\{S_0, S_1, \ldots, S_n\}$ of sets in $\mathfrak{S}$ such that $S = S_0 \subset S_1 \subset \cdots \subset S_n = T$ and $R_i = S_i - S_{i-1} \in \mathfrak{S}$ for $i = 1, \ldots, n$.

Lemma 2 will be applied to the following situation:

LEMMA 3. *The system $\mathfrak{S}$ of "intervals"*

$$[f, g) = \{(x, t) \in X \times \mathbb{R}: f(x) \leqslant t < g(x)\}, \quad g \geqslant f, \quad g, f \in F$$

is a semiring and

$$\nu\colon \mathfrak{S} \to \mathbb{R}_+, \quad \nu([f, g)) = \Lambda(g - f)$$

is well-defined and σ-additive.

Proof. $[f, g) = \Sigma_{n=1}^{\infty}[f_n, g_n)$ implies $[f(x), g(x)) = \Sigma_{n=1}^{\infty}[f_n(x), g_n(x))$ for every $x \in X$. From

Lemma 1 we get

$$g - f = \sum_{n=1}^{\infty} (g_n - f_n).$$

Let $h_n = g - f - \sum_{m=1}^{n}(g_m - f_m)$; then $h_n \downarrow 0$ so $\Lambda(h_n) \to 0$, i.e.,

$$\nu([f, g)) = \sum_{n=1}^{\infty} \nu([f_n, g_n)).$$

As in the case of "ordinary" intervals, the proof that $\mathfrak{S}$ is a semiring is an easy exercise.

Proof of the theorem. For ν as defined in Lemma 3, choose ρ according to Lemma 2. Let $0 \leqslant f \in F$. For $f_n = (n[f - f \wedge 1]) \wedge 1$ we have

(A) $$[0, \beta f_n) \uparrow \{f > 1\} \times [0, \beta) \quad \text{for every } \beta > 0.$$

Hence,

$$\mu(A) = \rho(A \times [0,1)), A \in \mathfrak{B}_F$$

is a well-defined measure with the property

$$\rho(\{f > 1\} \times [0, \beta)) = \beta\mu(\{f > 1\}), f \in F, \beta > 0.$$

Let $B^n = \sum_{i=1}^{n2^n}(C_i^n - C_{i+1}^n) \times [0, i2^{-n})$ with $C_i^n = \{i^{-1}2^n f > 1\}$ for $1 \leqslant i \leqslant n2^n$ and $C_i^n = \varnothing$ for $i = n2^n + 1$. Then, $B^n \uparrow [0, f)$ and we get

$$\Lambda(f) = \rho([0, f)) = \lim_{n \to \infty} \rho(B^n) = \int f\, d\mu.$$

Remarks.

(a) It is well known that one cannot dispense with "Stone's condition": $f \in F$ implies $f \wedge 1 \in F$. (Compare [**1**], 7XB for a counterexample.) However, any other condition which ensures for every $f \in F$ the existence of a sequence $f_n \in F$ with property (A) is also appropriate.

(b) Of course, the above theorem implies the Riesz representation theorem ([**3**], 56D). This is a well-known and immediate consequence of Dini's theorem. Another short proof for this special case has been given by Garling [**2**] (or [**4**], §22). Garling's method is closely related to Varadarajan's [**5**].

References

1. D. H. Fremlin, Topological Riesz Spaces and Measure Theory, Cambridge University Press, 1974.

2. D. J. H. Garling, A "short" proof of the Riesz representation theorem, Proc. Cambridge Philos. Soc., 73 (1973) 459–460.

3. P. R. Halmos, Measure Theory, Van Nostrand, Princeton, 1950.

4. R. B. Holmes, Geometric Functional Analysis and its Applications, Springer, New York-Heidelberg-Berlin, 1975.

5. V. S. Varadarajan, On a theorem of F. Riesz concerning the form of linear functionals, Fund. Math., XLVI (1958) 209–220.

ANSWER TO PHOTO ON PAGE 388

Hassler Whitney, one of the recipients of the 1982 Wolf Foundation Prize in Mathematics.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

AN ELEMENTARY APPROACH TO *NP*-COMPLETENESS

A. KEITH AUSTIN
Department of Pure Mathematics, The University, Sheffield S3 7RH, England

In 1971 Stephen Cook of the University of Toronto showed in [**1**] that any problem in NP, i.e., any problem which can be solved by a nondeterministic Turing machine in polynomial time, could be reduced to a problem of consistency in the propositional calculus. As a result of this work, the general consistency problem in the propositional calculus was said to be NP-complete.

Although this is a most significant result in computing science, it does, unfortunately, require an understanding of Turing machines to appreciate the general proof. However, the key idea of the proof can be illustrated by showing the reduction technique in a particular (but typical) case.

We will see how the following Traveling Salesman problem can be reduced to a problem in the propositional calculus. In the map, the vertices are towns and the lines are roads, each 10 miles long.

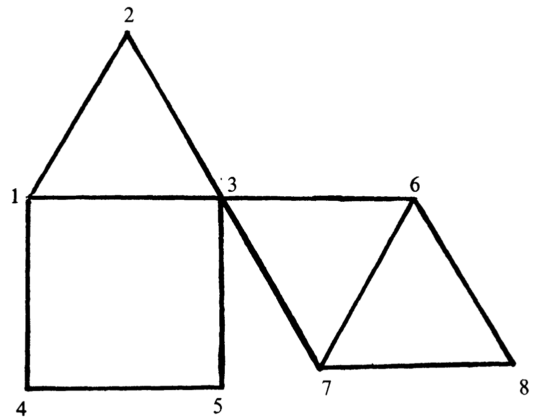


FIG. 1.

PROBLEM: Can the salesman start at 1 and visit all the towns in a journey of only 70 miles?

To express the problem in the propositional calculus, we write,

$$A_t^m = \text{after } 10 \times m \text{ miles he is at town } t,$$

for $m = 0, 1, 2, \ldots, 7$, $t = 1, 2, 3, \ldots, 8$. Given any journey of 70 miles, each A_t^m is either true or false.

We now express the conditions of the problem as logical statements.

If he is at 5 after 30 miles, then he is at 3 or 4 after 40 miles, i.e., if A_5^3 is true, then either A_3^4 or A_4^4 is true. We write this as

$$A_5^3 \to \left(A_3^4 \vee A_4^4\right),$$

writing $\to$ for "if-then" and $\vee$ for "or." Denote this expression by J_5^3. Similarly we have

$$A_5^m \to \left(A_3^{m+1} \vee A_4^{m+1}\right),$$

and

$$A_3^m \to \left(A_1^{m+1} \vee A_2^{m+1} \vee A_5^{m+1} \vee A_6^{m+1} \vee A_7^{m+1}\right),$$

for $m = 0, 1, 2, \ldots, 6$, and so on for each town. Denote each of these by the corresponding J_y^x. All these have to be true and so we write,

$$J = J_1^0 \,\&\, J_2^0 \,\&\cdots\&\, J_8^0 \,\&\, J_1^1 \,\&\, J_2^1 \,\&\cdots\&\, J_8^1 \,\&\cdots\&\, J_8^6,$$

writing & for "and."

Another condition is that each town has to be visited. That town 1 has to be visited can be expressed as

$$A_1^0 \vee A_1^1 \vee A_1^2 \vee \cdots \vee A_1^7$$

and similarly for the other towns. Let

$$V = \left(A_1^0 \vee \cdots \vee A_1^7\right) \& \left(A_2^0 \vee \cdots \vee A_2^7\right) \&\cdots\& \left(A_8^0 \vee \cdots \vee A_8^7\right).$$

Also he is only at one town at any one time, so we have, e.g., $A_3^5 \to \neg A_1^5$, writing $\neg$ for "not." Let

$$N_3^5 = \left(A_3^5 \to \neg A_1^5\right) \& \left(A_3^5 \to \neg A_2^5\right) \& \left(A_3^5 \to \neg A_4^5\right) \&\cdots\& \left(A_3^5 \to \neg A_8^5\right).$$

Let

$$N = N_1^0 \,\&\, N_2^0 \,\&\cdots\&\, N_8^7.$$

Finally, he has to start at 1 so we require A_1^0 to be true.

Thus, finding a route which satisfies all the conditions of the problem is equivalent to the propositional calculus problem of assigning true or false to each A_t^m so that the expression

$$E = J \,\&\, V \,\&\, N \,\&\, A_1^0$$

is true. This is a particular case of the general consistency problem for the propositional calculus. In the general consistency problem, an expression of the propositional calculus is given, and the aim is to assign true or false to the variables present so that the expression is true.

Cook's general proof follows similar lines; he describes the working of a nondeterministic Turing machine by means of an expression of the propositional calculus. The technique, in fact, goes back to Alan Turing's original paper of 1936, [**2**], in which he defined a Turing machine and described its working in terms of the predicate calculus, in order to show the undecidability of the predicate calculus.

When we have reduced the traveling salesman or other problems in *NP* to the propositional calculus, we are left with the question of whether this reduction helps us to solve these problems. The solution to the consistency problem which first comes to mind is that of simply trying all possible assignments of true or false to the A_t^m seriatim, and seeing if one makes E true. However there are 64 A_t^m and so the number of assignments is 2^{64}. Unfortunately, there is no known alternative method which reduces the length of the calculation from this exponential function of the number of variables to a polynomial function of the number of variables. Thus the present situation is that the reduction has not helped us with the solution of the original problem, as it is not difficult to find an exponential solution for the traveling salesman in its original form. The question is, whether the future will bring a polynomial solution to the general consistency problem, or a proof that no such solution exists.

References

1. S. A. Cook, The complexity of theorem proving procedures, Proc. 3rd Annual ACM Symposium on the Theory of Computing, 1971, 151–158.

2. A. M. Turing, On computable numbers, with an application to the Entscheidungsproblem, Proc. London Math. Soc., Ser. 2-42 (1936) 230–265.

positive integers, $n = n_1 + n_2 + n_3 + n_4$, $n_{i_1}n_{i_2} \neq n_{i_3}n_{i_4}$, for each permutation (i_1, i_2, i_3, i_4) of $(1, 2, 3, 4)$.

E 3006. *Proposed by J. R. Kuttler, Johns Hopkins University.*

Find a function of the form $\alpha(x) + \beta(y)$ which best approximates $\sin x \sin y$ on the square $S = \{(x, y)\colon 0 \leqslant x \leqslant 2\pi, 0 \leqslant y \leqslant 2\pi\}$ in the sup norm, i.e., so that

$$\sup_{(x,y)\in S} |\sin x \sin y - \alpha(x) - \beta(y)|$$

is as small as possible.

SOLUTIONS OF ELEMENTARY PROBLEMS

Sum of Squares of Digits

E 2925 [1982, 64]. *Proposed by T. Keller, Honolulu, HI.*

Let b be a positive integer. How many integers are there, each of which, when expressed in base b, is equal to the sum of the squares of its digits?

Solution by Edmund Butler, New Carrollton, MD. A simple calculation shows that 0 and 1 are the only 1-digit solutions, and there are no solutions with 3 or more digits. Let $N(b)$ be the total number of solutions and $N'(b)$ the number of 2-digit solutions, so $N(b) = N'(b) + 2$.

Then $N'(b)$ is the number of pairs (a_0, a_1) satisfying

$$(1) \qquad a_0 + a_1 b = a_0^2 + a_1^2 \qquad (0 \leqslant a_0 \leqslant b - 1, \quad 1 \leqslant a_1 \leqslant b - 1).$$

It is easily verified that there exists a 1-1 correspondence between pairs (a_0, a_1) satisfying (1) and pairs (p, q) satisfying

$$(2) \qquad p^2 + q^2 = 1 + b^2 \qquad (p \text{ odd}, 3 \leqslant p \leqslant b, \quad 1 \leqslant q \leqslant b - 1)$$

via the transformations $a_0 = (p + 1)/2$, $a_1 = (b + q)/2$.

Now we express the number of solutions to (2) in terms of the number of solutions $r(k)$ to

$$(3) \qquad c^2 + d^2 = k.$$

Suppose b is even. Then $1 + b^2$ is odd, so exactly one of p, q is odd. Thus given a solution (p, q) to (2) we can generate 3 others that solve (3) via $(c, d) = (-p, q)$; (q, p); and $(q, -p)$. We also add the 8 remaining solutions $(\pm 1, \pm b)$; $(\pm b, \pm 1)$. This shows that $r(1 + b^2) = 4N'(b) + 8 = 4N(b)$.

Now suppose b is odd. Then $1 + b^2 \equiv 2 \pmod 4$; hence both p and q must be odd. Thus from any solution (p, q) to (2) we can generate another solution to (3) via $(-p, q)$. We also add the 4 remaining uncounted solutions $(\pm 1, \pm b)$. This shows

$$r(1 + b^2) = 2N'(b) + 4 = 2N(b).$$

Now $r(k)$ is a famous quantity and can be computed from a well-known formula (for example, Hardy and Wright, *Introduction to the Theory of Numbers*, 241). Using the fact that no prime of the form $4j + 3$ can divide $1 + b^2$, we find that

$$r(1 + b^2) = \begin{cases} 4\tau(1 + b^2), & b \text{ even} \\ 2\tau(1 + b^2), & b \text{ odd} \end{cases}$$

where $\tau(n)$ is the number of positive integral divisors of n. Therefore we find $N(b) = \tau(1 + b^2)$.

Also solved by M. Golomb, Purdue University; R. T. Koether, Hampden-Sydney, VA; O. P. Lossers, Netherlands; G. Shulman, Teaneck, NJ, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by November 30, 1983. The solver's full post-office address should be on each sheet.

6430. *Proposed by J. Borwein, Carnegie-Mellon University.*

Recall that the p-trace average of a positive definite symmetric $n \times n$ matrix C (we write $C > 0$) is given by

$$j_p(C) := \left(\frac{1}{n}\sum_{i=1}^{n}\lambda_i^p\right)^{1/p}$$

for $-\infty < p < \infty$, $p \neq 0$. Here $\lambda_1 < \lambda_2 \leqslant \cdots \leqslant \lambda_n$ are the eigenvalues of C. Extend this definition by continuity for $p = 0, \infty$, and $-\infty$. Thus

$$j_0(C) := (\det C)^{1/n},\ j_\infty(C) := \lambda_n, \quad \text{and} \quad j_{-\infty}(C) := \lambda_1.$$

Let $\Delta(C)$ be the diagonal matrix whose diagonal coincides with the diagonal of C. Show that

(i) $j_p(C) \leqslant j_p(\Delta(C))$ for $-\infty \leqslant p \leqslant 1$,

and

(ii) $j_p(C) \geqslant j_p(\Delta(C))$ for $1 \leqslant p \leqslant \infty$.

Note that for $p = 0$ this is the well-known Hadamard inequality.

6431. *Proposed by Gabor J. Szekely and Andras Zempleni, Budapest, Hungary.*

If X and Y are independent nonnegative integer-valued random variables and XY has a Poisson distribution, show that either X or Y takes at most two values, zero and one, with probability one. (In this sense the Poisson distribution is irreducible.)

6432. *Proposed by Michael Barr, McGill University.*

Problem 4 on p. 348 of Jacobson's *Basic Algebra I* reads, "Call a linear transformation *normal* if it commutes with its adjoint [with respect to a symmetric or alternating, nonsingular bilinear form on the vector space V—*not* its classical adjoint]. ... Show that if U [a subspace of V] is stabilized by a normal linear transformation T, then $U^\perp$ is stabilized by T."

(a) Show by example that the assertion is false.

(b) Show that if the form is anisotropic on a finite dimensional space (no nonzero vector is orthogonal to itself—this requires the form to be symmetric) and the ground field is perfect, then T is diagonalizable over the algebraic closure of the ground field.

(c)* Is it the case that if U is stabilized by a linear transformation T normal with respect to an anisotropic form (over a perfect field, if necessary), then $U^\perp$ is stabilized by T?

6433. *Proposed by Edmund Butler, New Carrollton, MD.*

Let $X_1, X_2, \ldots, X_{n+1}$ be a sequence of independent random variables uniformly distributed on $[0, 1]$. For any sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ of ± 1's let

$$P_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \Pr\{\operatorname{sign}(x_1 - x_2) = \varepsilon_1, \operatorname{sign}(x_2 - x_3) = \varepsilon_2, \ldots, \operatorname{sign}(x_{n+1} - x_n) = \varepsilon_n\}.$$

(a) Show that $M_n = \operatorname{Max}_{\varepsilon_j = \pm 1} P_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is assumed when the ε's alternate in sign.

(b) Find $\lim_{n\to\infty} M_n^{1/n}$.

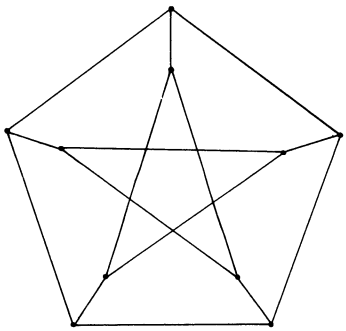
6434. *Proposed by Allen J. Schwenk, University of Waterloo.*

The Petersen graph shown below has 15 edges. Its adjacency matrix P is formed by setting the (i, j) entry equal to 1 if vertex i is adjacent to vertex j and to 0 otherwise.

(a) Show that its characteristic polynomial is

$$(x-3)(x-1)^5(x+2)^4.$$

(b) Can the 45 edges of the complete graph K_{10} be partitioned into three copies of the Petersen graph?



6435. *Proposed by Thomas Q. Sibley, Cuttington University College, Liberia.*

Let m be any finitely additive extension of Lebesgue measure to all subsets of $[0,1]$. Are there two subsets A, B of $[0,1]$ such that $A \cap B = \varnothing$, $A \cup B = [0,1]$ and for any $0 \leqslant a < b \leqslant 1$

$$m(A \cap [a,b]) = m(B \cap [a,b]) = \frac{1}{2}(b-a)?$$

SOLUTIONS OF ADVANCED PROBLEMS

Volume of a Certain Convex Polytope

5872* [1972, 913]. *Proposed by Shmuel Schreiber, Bal-Ilan University, Israel.*

Let C_n denote the region in Euclidean x space defined by $x_i \geqslant 0$ for $i = 1,\ldots, n$ and $y_i \geqslant 0$ for $i = 1,\ldots, n$, where

$$\begin{aligned} x_1 &= 1 - 2y_1 + y_2 \\ x_i &= 1 + y_{i-1} - 2y_i + y_{i+1} (2 \leqslant i \leqslant n-1) \\ x_n &= 1 + y_{n-1} - 2y_n. \end{aligned}$$

Prove that C_n is a convex polytope of the combinatorial type of a cube and that its volume is $(n+1)^{n-1}/(n!)$. (The result has some use in tournament theory.)

Solution by H. Debrunner, Universität Bern, Switzerland.

1. In contrast to the problem text we prefer to consider C_n in the euclidean space E^n given by the n-tuples $y = (y_1,\ldots, y_n)$. The combinatorial type is invariant under the affine mapping χ relating x to y, but the volume will differ by a factor $n+1$, the absolute value of det χ (cf. [3]). The problem shows only part of cyclic symmetry; in order to complete it we introduce a further coordinate y_{n+1} which takes the value 0 on C_n. Consequently, we put $m := n+1$ and we will prove that C_{m-1} has the combinatorial type of an $(m-1)$-cube and has in its y-space $(m-1)$-volume vol $C_{m-1} = m^{m-3}/(m-1)!$. We assume $m \geqslant 3$ throughout.

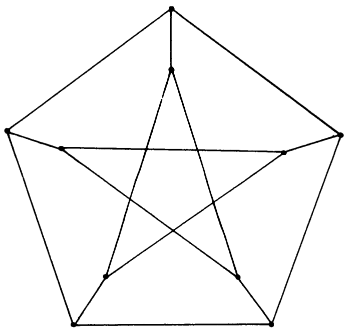
6434. *Proposed by Allen J. Schwenk, University of Waterloo.*

The Petersen graph shown below has 15 edges. Its adjacency matrix P is formed by setting the (i, j) entry equal to 1 if vertex i is adjacent to vertex j and to 0 otherwise.

(a) Show that its characteristic polynomial is

$$(x-3)(x-1)^5(x+2)^4.$$

(b) Can the 45 edges of the complete graph K_{10} be partitioned into three copies of the Petersen graph?



6435. *Proposed by Thomas Q. Sibley, Cuttington University College, Liberia.*

Let m be any finitely additive extension of Lebesgue measure to all subsets of $[0,1]$. Are there two subsets A, B of $[0,1]$ such that $A \cap B = \varnothing, A \cup B = [0,1]$ and for any $0 \leqslant a < b \leqslant 1$

$$m(A \cap [a,b]) = m(B \cap [a,b]) = \frac{1}{2}(b-a)?$$

SOLUTIONS OF ADVANCED PROBLEMS

Volume of a Certain Convex Polytope

5872* [1972, 913]. *Proposed by Shmuel Schreiber, Bal-Ilan University, Israel.*

Let C_n denote the region in Euclidean x space defined by $x_i \geqslant 0$ for $i = 1, \ldots, n$ and $y_i \geqslant 0$ for $i = 1, \ldots, n$, where

$$x_1 = 1 - 2y_1 + y_2$$
$$x_i = 1 + y_{i-1} - 2y_i + y_{i+1} (2 \leqslant i \leqslant n-1)$$
$$x_n = 1 + y_{n-1} - 2y_n.$$

Prove that C_n is a convex polytope of the combinatorial type of a cube and that its volume is $(n+1)^{n-1}/(n!)$. (The result has some use in tournament theory.)

Solution by H. Debrunner, Universität Bern, Switzerland.

1. In contrast to the problem text we prefer to consider C_n in the euclidean space E^n given by the n-tuples $y = (y_1, \ldots, y_n)$. The combinatorial type is invariant under the affine mapping χ relating x to y, but the volume will differ by a factor $n+1$, the absolute value of det χ (cf. [3]). The problem shows only part of cyclic symmetry; in order to complete it we introduce a further coordinate y_{n+1} which takes the value 0 on C_n. Consequently, we put $m := n+1$ and we will prove that C_{m-1} has the combinatorial type of an $(m-1)$-cube and has in its y-space $(m-1)$-volume vol $C_{m-1} = m^{m-3}/(m-1)!$. We assume $m \geqslant 3$ throughout.

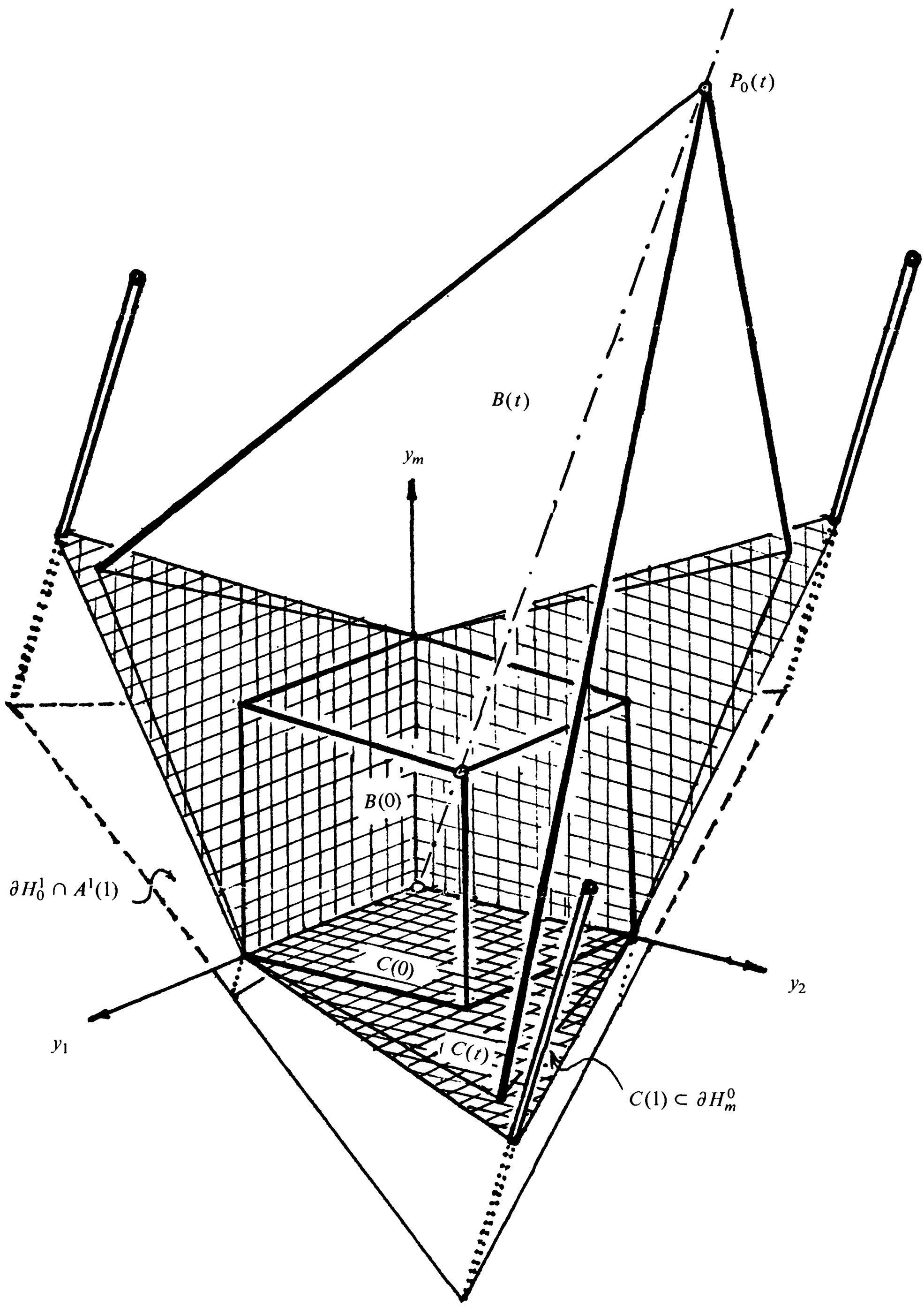

$P_0(t)$
$B(t)$
y_m
$B(0)$
$\partial H_0^1 \cap A^1(1)$
$C(0)$
y_2
y_1
$C(t)$
$C(1) \subset \partial H_m^0$

FIG. 1.

2. For $t \in \mathbb{R}$, let $F_m(t)$ denote the $m \times m$-matrix with elements $f_{ij} = 2, f_{ij} = -t, f_{ij} = 0$ in the cases $j = i, j = i \pm 1, j \neq i, i \pm 1$, resp., where indices are taken modulo m so that we have cyclic symmetry. For $i = 1, \dots, m$, we let $f_i(t) \in E^m$ denote the ith row of $F_m(t)$ and $e_i \in E^m$ the vector with coordinates δ_{ij} (Kronecker symbol). We shall consider the halfspaces $H_i^0(t)$ and $H_i^1(t)$ of E^m given by $y \cdot e_i \geqslant 0$ and $y \cdot f_i(t) \leqslant 1$, resp., and also their boundary hyperplanes $\partial H_i^0(t)$, $\partial H_i^1(t)$ given by the equations

$$\binom{0}{i}: \quad y \cdot e_i = 0, \quad \binom{1}{i}: \quad y \cdot f_i(t) - 1 = 0 \quad (i = 1, \dots, m).$$

We introduce the (unbounded) polytopes $A^0(t) := H_1^0(t) \cap \cdots \cap H_m^0(t)$ and $A^1(t) := H_1^1(t) \cap \cdots \cap H_m^1(t)$ and their intersection $B(t) := A^0(t) \cap A^1(t)$. Of course $H_i^0(t)$, $\partial H_1^0(t)$ and $A^0(t)$ are actually independent of t so that we may also use the shorter notations H_i^0, ∂H_i^0, A^0.

3. Two values of t merit special interest: for $t = 0$, $B(0)$ is an honest metric m-cube of side length $1/2$; and for $t = 1$, the C_{m-1} of the problem text is the facet $y_m = 0$ of the polytope $B(1)$. So we introduce this facet also in the general case by $C(t) := \partial H_m^0 \cap B(t)$. (See Fig. 1.) Then $C(0)$ is an $(m-1)$-cube and $C(1)$ must be shown to have the same combinatorial type. Our method to tackle this problem consists in proving that the combinatorial types of $B(t)$ and $C(t)$ remain unchanged when t varies from 0 to 1. Actually we will meet the difficulty that the combinatorial type of $B(t)$ changes exactly for $t = 1$, since the vertex $(2 - 2t)^{-1}(e_1 + \cdots + e_m)$ of $B(t)$, diagonally opposite to the origin, tends to infinity as $t \to 1-$. But we will see that this does not affect the type of the facet $C(t)$ and, on the other hand, will help to determine vol C_{m-1}.

4. We will need various determinants related to $F_m(t)$; for any square matrix P we denote by P^* its determinant. Let $X_j(t)$ for $j = 1, \dots, m-1$ denote the principal $j \times j$-minor of $F_m(t)$; note that $-X_n(1)$ is the matrix of coefficients appearing in the original problem text. If in $X_j(t)$ we replace the first column by $(1, \dots, 1) \in E^j$, the resulting matrix will be denoted by $Y_j(t)$; here we remark that the determinant value would be the same (also in sign) if we had replaced the last column of $X_j(t)$ by $(1, \dots, 1)$. Now, expanding along the first row and then some of the minors along their first columns, we get recursion formulas

$$X_j^*(t) = 2X_{j-1}^*(t) - t^2 X_{j-2}^*(t) \tag{1}$$

and

$$Y_j^*(t) = X_{j-1}^*(t) + tY_{j-1}^*(t). \tag{2}$$

With the starting values $X_0^*(t) = 1$, $X_1^*(t) = 2$, $X_2^*(t) = 4 - t^2$, $Y_1^*(t) = 1$, $Y_2^*(t) = 2 + t$, induction arguments now yield the special values

$$X_n^*(1) = n + 1, \quad Y_n^*(1) = \tfrac{1}{2}n(n+1) \quad (n = 1, 2, \dots) \tag{3}$$

and the inequalities

$$X_{n+1}^*(t) > X_n^*(t) > 0 \quad \text{and} \quad Y_n^*(t) > 0 \quad \text{for} \quad 0 \leqslant t \leqslant 1. \tag{4}$$

Since $F_m(t)$ is a circulant matrix, its determinant can easily be determined explicitly (as the product of its eigenvalues), cf. [**1**, p. 73] or [**2**], namely

$$F_m^*(t) = 2^{m-1}(2 - 2t) \prod_{j=1}^{m-1} \left(1 - t\cos\frac{2j\pi}{m}\right). \tag{5}$$

Finally, we consider the symmetric matrix $Z_{m+1}(t)$ resulting from $F_m(t)$ by bordering with a foremost row and column $f_0 := (0, 1, \dots, 1) \in E^{m+1}$ intersecting in the common entry 0. If $t \neq 1$, we multiply the sum of the last m columns of $Z_m(t)$ by $(2 - 2t)^{-1}$ and subtract this from the foremost column; so we get

$$Z_{m+1}^*(t) = -(2 - 2t)^{-1} m F_m^*(t) \quad \text{for} \quad t \neq 1. \tag{6}$$

If $t = 1$, we add the last $m - 1$ columns of $Z_{m+1}(1)$ to the column $(1, f_1(1))$ and similarly for the rows, and, in view of $\Sigma_1^m f_i(1) = 0$, we get

$$Z_{m+1}^*(1) = -m^2 X_{m-1}^*(1) = -m^3. \tag{7}$$

5. Obviously our A^0 is a simplicial cone with apex 0. Next we show that $A^1(t)$ is, for $0 \leqslant t < 1$, a simplicial cone with apex $p_0 := (2 - 2t)^{-1}(e_1 + \cdots + e_m)$ and, for $t = 1$, a prismatic region with simplicial crossection and with axis $\mathbb{R}(e_1 + \cdots + e_m)$. This is clear from the inequalities defining the $H_i^1(t)$ $(i = 1, \ldots, m)$ once we know that $f_1(t), \ldots, f_m(t)$ are linearly independent for $0 \leqslant t < 1$ and span an $(m-1)$-dimensional subspace orthogonal to $e_1 + \cdots + e_m$ for $t = 1$. The orthogonality conditions $f_i(1) \cdot (e_1 + \cdots + e_m) = 0$ are easily checked; the remaining assertion follows from rank considerations for $F_m(t)$. In fact, (5) shows $\operatorname{rank}(F_m(t)) = m$ for $0 \leqslant t < 1$, and (5) together with (4) shows $\operatorname{rank}(F_m(1)) = m - 1$, since $X_{m-1}(1)$ is a minor of $F_m(1)$.

6. As t runs from 0 upwards, the combinatorial type of $B(t_0)$ is the same as that of the cube $B(0)$, if for each $t \in [0, t_0]$

(i) the m hyperplanes $\partial H_1^{\varepsilon_1}(t), \ldots, \partial H_m^{\varepsilon_m}(t)$ (where $\varepsilon_i \in \{0, 1\}$ for $i = 1, \ldots, m$) have linearly independent normal vectors and thus intersect in exactly one point $P^\varepsilon(t)$; this for each of the 2^m possible choices of $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$. And in addition:

(ii) the $m + 1$ hyperplanes $\partial H_i^{1-\varepsilon_i}(t)$, $\partial H_1^{\varepsilon_1}(t), \ldots, \partial H_m^{\varepsilon_m}(t)$ (where $i \in \{1, \ldots, m\}$ and $\varepsilon_j \in \{0, 1\}$) have no common point; this for each of the $2^{m-1}m$ different choices for i and ε_j $(j \neq i)$. By continuity the condition (ii) guarantees in the presence of (i) that each point $P^\varepsilon(t)$ keeps in the interior of the halfspaces $H_i^{1-\varepsilon_i}(t)$, since for $t = 0$ it is in this interior. So $P^\varepsilon(t_0)$ will be a vertex of $B(t_0)$, and for $t = t_0$ we will have the same incidence relations between vertices and facets of $B(t)$ as for $t = 0$. For the sufficiency of this condition see [**3**, p. 41, Exercise 3]. The same conditions (i) and (ii), restricted to the fixed choice $\varepsilon_m = 0$, will guarantee that the combinatorial type of $C(t_0)$ $(= \partial H_m^0(t_0) \cap B(t_0))$ is the same as that of the $(m-1)$-cube $C(0)$. For the $(m-2)$-dimensional hyperplanes $\partial H_i^{\varepsilon_i}(t) \cap \partial H_m^0$ $(i = 1, \ldots, m-1)$ in the $(m-1)$-plane ∂H_m^0 intersect just like the hyperplanes listed in (i) if $\varepsilon_m = 0$.

7. We have already checked in §5 that condition (i), for $\varepsilon = (1, 1, \ldots, 1)$, is true if $0 \leqslant t < 1$ and fails if $t = 1$. We next check that (i) is true in the case $\Sigma_1^m \varepsilon_i < m$ if $0 \leqslant t \leqslant 1$. Considering the cyclic symmetry of $F_m(t)$, we may assume without loss of generality that one of the vanishing ε_i's is $\varepsilon_1 = 0$. The matrix $N_m^\varepsilon(t)$ which checks the linear independence of the normals to the hyperplanes listed in (i) arises from $F_m(t)$ by replacing the ith row by e_i if $\varepsilon_i = 0$ $(i = 1, \ldots, m)$. If we assume $\varepsilon_i = 0$ exactly for the index values $1 = i_1 < i_2 < \cdots < i_r$, then with $i_{r+1} := m + 1$, we easily see

$$\det N_m^\varepsilon(t) = \prod_{s=1}^{r} X_{i_{s+1}-i_s-1}^*(t),$$

so by (5) $N_m^\varepsilon(t)$ has rank m for $0 \leqslant t \leqslant 1$. Summing up, condition (i) is generally true for $0 \leqslant t < 1$, and in the restricted case $\varepsilon_m = 0$ it is true for $0 \leqslant t \leqslant 1$.

8. To deal with condition (ii) we again exploit cyclic symmetry by assuming without loss of generality that $i = 1$. The listed hyperplanes then have no point in common if the system of $m + 1$ equations $\binom{0}{1}, \binom{1}{1}, \binom{\varepsilon_2}{2}, \ldots, \binom{\varepsilon_m}{m}$ (for notation see §2) has no solution $y \in E^m$. It amounts to the same to say that this system made homogeneous (by writing, e.g., first $-y_i/y_0$ in place of y_i $(i = 1, \ldots, m)$ and then multiplying the equations by y_0) has only the trivial solution $(0, \ldots, 0) \in E^{m+1}$. The $(m+1) \times (m+1)$-matrix $M^\varepsilon(t)$ of this system of equations can be got from the matrix $F_m(t)$ by adjoining first a foremost row $e_1 = (1, 0, \ldots, 0) \in E^m$, then replacing the ith row

$f_i(t)$ by e_i if $\varepsilon_i = 0$ $(i = 2,\dots, m)$, and finally adjoining a foremost column $(0, 1, \varepsilon_2, \varepsilon_3,\dots, \varepsilon_m)$. For the computation of $|\det M^\varepsilon(t)|$ we may first delete all the rows of $M^\varepsilon(t)$ beginning with an entry 0 and the corresponding columns which cross these rows in their only nonvanishing entry 1. If then we expand along the first remaining row, which will be $(1, -t, 0,\dots, 0, 0)$ or $(1, -t, 0,\dots, 0, -t)$ according to whether $\varepsilon_m = 0$ or $\varepsilon_m = 1$, we see (in the second case after some reordering of the columns of the algebraic complement to the last entry of the first line) that $\det M^\varepsilon(t)$ is a sum of products consisting of factors t^j, $X_r^*(t)$ and $Y_s^*(t)$ $(j, r, s \in \{1, 2,\dots\})$, all the summands being of the same sign. So, in view of (5), we have $\det M^\varepsilon(t) \neq 0$ for $0 \leqslant t \leqslant 1$. Summing up, condition (ii) is true for $0 \leqslant t \leqslant 1$ in the general case as well as in the restricted case $\varepsilon_m = 0$. Accordingly, we now have proved that C_{m-1} has the combinatorial type of an $(m-1)$-cube and, moreover, that the same holds for $C(t)$ $(0 \leqslant t \leqslant 1)$ and that $B(t)$ has the combinatorial type of an m-cube $(0 \leqslant t < 1)$.

9. To compute the $(m-1)$-volume of the facet $C(1)$ we introduce a further halfspace

$$H_0^1 = \left\{ y \in E^m \middle| y \cdot \sum_1^m e_i \geqslant 0 \right\}.$$

Its boundary hyperplane ∂H_0^1 contains the origin, has normal vector $(1, 1,\dots, 1)$ and is the subspace spanned by $f_1(1),\dots, f_m(1)$. Each of the sets $B(t)$, $A^0(t)$, $A^1(t)$, $\{0\}$, H_0^1 is invariant under the orthogonal transformation ζ which permutes the coordinate axes cyclically. So we have the same cyclic symmetry on $\partial A^0(1) \cap A^1(1)$ which consists of the m congruent facets $C(1), \zeta C(1),\dots, \zeta^{m-1}C(1)$ of $B(1)$. But orthogonal projection onto ∂H_0^1 (along the axis $\mathbb{R}(e_1 + \cdots + e_m)$ of the prism $A^1(1)$) maps $\partial A^0(1) \cap A^1(1)$ bijectively onto $\partial H_0^1 \cap A^1(1)$ and maps each of the facets $\zeta^i C(1)$ with projection factor $m^{-1/2}$ $(= \cos\measuredangle(e_i, \Sigma e_j))$. (See Fig. 1 and note that the prism $A^1(1)$, indicated there by stakes on the edges, has nonregular cross section for $m > 3$.) So if, for $0 \leqslant t \leqslant 1$, we define $\gamma_0(t)$ as the $(m-1)$-volume of $\partial H_0^1 \cap A^1(t)$, we have

$$\gamma_0(1) = m \cdot m^{-1/2} \operatorname{vol} C(1) = m^{1/2} \operatorname{vol} C_{m-1}. \tag{8}$$

To evaluate the left-hand side of (8) by means of $\gamma_0(1) = \lim_{t \to 1-} \gamma_0(t)$, we consider, for $0 \leqslant t < 1$, the m-simplex $S(t) = H_0^1 \cap A^1(t)$. It has the apex $p_0 = (2 - 2t)^{-1}\Sigma e_j$ of $A^1(t)$ as a vertex, opposite to the facet $\partial H_0^1 \cap A^1(t)$, and the corresponding height, which coincides by ζ-symmetry with the distance from O to p_0, is

$$h_0 = (2 - 2t)^{-1} m^{1/2}. \tag{9}$$

We have for the m-volume $V(t)$ of $S(t)$ the equation

$$mV(t) = \gamma_0(t) h_0. \tag{10}$$

For $i = 1,\dots, m$, let $p_i \in \partial H_0^1$ be the vertex of $S(t)$ which is opposite to the ith facet $\partial H_i^1(t) \cap S(t)$ having outer normal vector $f_i(t)$, and let $\gamma_i(t)$ denote the $(m-1)$-volume of this facet, h_i the distance from p_i to $\partial H_i^1(t)$ (ith height) and d_i the distance from the origin O to $\partial H_i^1(t)$. We dissect the simplex $S(t)$ into m pyramids each having the origin O (which is a relative interior point of the facet $\partial H_0^1 \cap S(t)$) as its apex and a facet of $S(t)$ as its base. All these pyramids are congruent by means of ζ. So we have

$$mV(t) = \gamma_i(t) h_i = \sum_{j=1}^m \gamma_j(t) d_j = m\gamma_i(t) d_i \quad (i = 1,\dots, m)$$

and, in particular,

$$h_i / d_i = m \quad (i = 1,\dots, m). \tag{11}$$

If we evaluate the left-hand side of equation $\binom{1}{j}$ for the vertex p_i, Hesse's normal form formula yields

$$p_i \cdot f_j(t) - 1 = \pm\delta_{ij}\frac{h_i}{d_i} = \pm\delta_{ij}m.$$

All these incidence relations between vertices and facets of $S(t)$ are contained in the matrix equation

$$(12)\qquad \begin{bmatrix} 1 & p_0 \\ \vdots & \vdots \\ 1 & p_m \end{bmatrix} \begin{bmatrix} 0 & -1 & \cdots & -1 \\ 1 & & & \\ \vdots & f_1(t) & \cdots & f_m(t) \\ 1 & & & \end{bmatrix} = \begin{bmatrix} h_0 m^{1/2} & & & 0 \\ & \pm m & & \\ & & \ddots & \\ 0 & & & \pm m \end{bmatrix}$$

It is well known that the determinant of the first matrix is $m!V(t)$. That of the second is $-Z^*_{m+1}(t)$, see §4. So, with (10) and (6), the determinant relation associated to (12) reads

$$(m-1)!\gamma_0(t)h_0Z^*_{m+1}(t) = \pm h_0 m^{m+1/2}$$

or

$$(13)\qquad \gamma_0(t) = |m^{m+1/2}/(m-1)!Z^*_{m+1}(t)|.$$

Combining this with (7) and (8), we get

$$\operatorname{vol} C_{m-1} = m^{-1/2}\left|\lim_{t\to 1-}\gamma_0(t)\right| = |m^m/(m-1)!Z^*_{m+1}(1)| = m^{m-3}/(m-1)!,$$

as asserted.

References

1. Ph.J. Davis, Circulant Matrices, Wiley, New York, 1979.
2. E. T. Wong, Polygons, circulant matrices, and Moore-Penrose inverses, this MONTHLY, 88(1981) 509–515.
3. B. Grunbaum, Convex Polytopes, Wiley, London, 1967.

Images of Monotone Functions

6218 [1978, 500; 1982, 134]. *Proposed by M. J. Pelling, Balliol College, Oxford, England.*

Let S be a subset of the real line R having cardinality of the continuum. Is there always a monotonic $f\colon R \to R$ such that $m^*f(S) > 0$ where m^* is outer Lebesgue measure?

Solution by A. W. Miller, University of Texas. This was partially solved by Fred Galvin (see this MONTHLY, 89 (1982) 134–135). He showed that the continuum hypothesis (actually SBCT which is much weaker) implies that there is a subset $S \subseteq R$ of cardinality the continuum such that every monotonic image of S has measure zero.

A positive answer to Pelling's problem is also consistent with the ZFC (Zermelo-Fraenkel set theory with the axiom of choice). A set of reals X has universal measure zero iff for every countably additive, nonatomic measure μ on the reals, $\mu(X) = 0$.

$$p_i \cdot f_j(t) - 1 = \pm\delta_{ij}\frac{h_i}{d_i} = \pm\delta_{ij} m.$$

All these incidence relations between vertices and facets of $S(t)$ are contained in the matrix equation

$$\begin{bmatrix} 1 & p_0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & p_m \end{bmatrix} \begin{bmatrix} 0 & -1 & \cdots & -1 \\ 1 & & & \\ \vdots & f_1(t) & \cdots & f_m(t) \\ 1 & & & \end{bmatrix}$$

$$= \begin{bmatrix} h_0 m^{1/2} & & & 0 \\ & \pm m & & \\ & & \ddots & \\ 0 & & & \pm m \end{bmatrix} \tag{12}$$

It is well known that the determinant of the first matrix is $m!V(t)$. That of the second is $-Z^*_{m+1}(t)$, see §4. So, with (10) and (6), the determinant relation associated to (12) reads

$$(m-1)!\gamma_0(t) h_0 Z^*_{m+1}(t) = \pm h_0 m^{m+1/2}$$

or

$$\gamma_0(t) = |m^{m+1/2}/(m-1)!Z^*_{m+1}(t)|. \tag{13}$$

Combining this with (7) and (8), we get

$$\operatorname{vol} C_{m-1} = m^{-1/2}\left|\lim_{t\to 1-} \gamma_0(t)\right| = |m^m/(m-1)!Z^*_{m+1}(1)| = m^{m-3}/(m-1)!,$$

as asserted.

References

1. Ph.J. Davis, Circulant Matrices, Wiley, New York, 1979.
2. E. T. Wong, Polygons, circulant matrices, and Moore-Penrose inverses, this MONTHLY, 88(1981) 509–515.
3. B. Grunbaum, Convex Polytopes, Wiley, London, 1967.

Images of Monotone Functions

6218 [1978, 500; 1982, 134]. *Proposed by M. J. Pelling, Balliol College, Oxford, England.*

Let S be a subset of the real line R having cardinality of the continuum. Is there always a monotonic f: $R \to R$ such that $m^* f(S) > 0$ where m^* is outer Lebesgue measure?

Solution by A. W. Miller, University of Texas. This was partially solved by Fred Galvin (see this MONTHLY, 89 (1982) 134–135). He showed that the continuum hypothesis (actually SBCT which is much weaker) implies that there is a subset $S \subseteq R$ of cardinality the continuum such that every monotonic image of S has measure zero.

A positive answer to Pelling's problem is also consistent with the ZFC (Zermelo-Fraenkel set theory with the axiom of choice). A set of reals X has universal measure zero iff for every countably additive, nonatomic measure μ on the reals, $\mu(X) = 0$.

It is an unpublished result of J. Baumgartner and R. Laver that if ZFC is consistent, then so is ZFC plus every set of reals of cardinality the continuum that fails to have universal measure zero. This result was announced in Laver, R. "On the consistency of Borel's conjecture," Acta Math., 137 (1976) 151–169. A different proof of it will also appear in Miller, A., "Mapping a set of reals onto the reals," to appear in the Journal of Symbolic Logic (1983). This result combined with the following construction of Szpilrajn-Marczewski completely settles Pelling's problem.

THEOREM (Szpilrajn-Marczewski). *Suppose $S \subseteq R$ and μ is a countable additive, nonatomic measure such that $\mu(S) > 0$. Then there exists a monotonic function $f\colon R \to R$ such that $m^*(f(s)) >$ 0.*

Proof. We may assume that μ does not vanish on any interval since we could replace μ by $\frac{1}{2}(\mu + \lambda)$ where λ is Lebesgue measure. Now define f by

$$f(x) = \mu([0, x]).$$

Since μ vanishes on no intervals and is atomless, we see that f is strictly increasing and continuous and thus a homeomorphism. Define

$$\nu(B) = \mu(f^{-1}(B))$$

for any Borel set B. We are done if we show that ν is Lebesgue measure. Suppose $I = [a, b]$ is any interval. Then

$$f^{-1}(I) = \{y\colon \mu([0, y]) \in I\} = [c, d]$$

where $\mu([0, c]) = a$ and $\mu([0, d]) = b$. But then

$$\nu(I) = \mu(f^{-1}(I)) = \mu([c, d]) = b - a.$$

Since ν agrees with Lebesgue measure on the intervals and the intervals generate the Borel sets, we have that ν is Lebesgue measure. □

Stochastic Matrices

6366 [1981, 711]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a row stochastic matrix such that $\|A\| = 1$, $\|\ \|$ being the operator norm induced by the Euclidean vector norm. Show that A is doubly stochastic.

Solution by Enzo R. Gentile, Universidad de Buenos Aires, Argentina. Let $A = (a_{ij})$ be a real $n \times n$ matrix. We shall prove that

$$\|A\| = \max_{x \neq 0} \frac{|Ax|}{|x|} = 1 \quad \text{and} \quad \sum_{i,j} a_{ij} = n$$

imply that A is doubly stochastic. Since $\|A\| = \|A^T\| = 1$ where A^T is the transpose of A, it will be enough to prove that the sum

$$\begin{aligned} S &= \sum_j \Big(1 - \sum_i a_{ij}\Big)^2 \\ &= \sum_j \Big(1 + \Big(\sum_i a_{ij}\Big)^2 - 2\Big(\sum_i a_{ij}\Big)\Big) \\ &= -n + \sum_j \Big(\sum_i a_{ij}\Big)^2 \end{aligned}$$

is 0.

It is an unpublished result of J. Baumgartner and R. Laver that if ZFC is consistent, then so is ZFC plus every set of reals of cardinality the continuum that fails to have universal measure zero. This result was announced in Laver, R. "On the consistency of Borel's conjecture," Acta Math., 137 (1976) 151–169. A different proof of it will also appear in Miller, A., "Mapping a set of reals onto the reals," to appear in the Journal of Symbolic Logic (1983). This result combined with the following construction of Szpilrajn-Marczewski completely settles Pelling's problem.

THEOREM (Szpilrajn-Marczewski). *Suppose $S \subseteq R$ and μ is a countable additive, nonatomic measure such that $\mu(S) > 0$. Then there exists a monotonic function $f\colon R \to R$ such that $m^*(f(s)) > 0$.*

Proof. We may assume that μ does not vanish on any interval since we could replace μ by $\frac{1}{2}(\mu + \lambda)$ where λ is Lebesgue measure. Now define f by

$$f(x) = \mu([0, x]).$$

Since μ vanishes on no intervals and is atomless, we see that f is strictly increasing and continuous and thus a homeomorphism. Define

$$\nu(B) = \mu(f^{-1}(B))$$

for any Borel set B. We are done if we show that ν is Lebesgue measure. Suppose $I = [a, b]$ is any interval. Then

$$f^{-1}(I) = \{y\colon \mu([0, y]) \in I\} = [c, d]$$

where $\mu([0, c]) = a$ and $\mu([0, d]) = b$. But then

$$\nu(I) = \mu(f^{-1}(I)) = \mu([c, d]) = b - a.$$

Since ν agrees with Lebesgue measure on the intervals and the intervals generate the Borel sets, we have that ν is Lebesgue measure. □

Stochastic Matrices

6366 [1981, 711]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a row stochastic matrix such that $\|A\| = 1$, $\|\ \|$ being the operator norm induced by the Euclidean vector norm. Show that A is doubly stochastic.

Solution by Enzo R. Gentile, Universidad de Buenos Aires, Argentina. Let $A = (a_{ij})$ be a real $n \times n$ matrix. We shall prove that

$$\|A\| = \max_{x \neq 0} \frac{|Ax|}{|x|} = 1 \quad \text{and} \quad \sum_{i,j} a_{ij} = n$$

imply that A is doubly stochastic. Since $\|A\| = \|A^T\| = 1$ where A^T is the transpose of A, it will be enough to prove that the sum

$$\begin{aligned} S &= \sum_j \Big(1 - \sum_i a_{ij}\Big)^2 \\ &= \sum_j \Big(1 + \Big(\sum_i a_{ij}\Big)^2 - 2\Big(\sum_i a_{ij}\Big)\Big) \\ &= -n + \sum_j \Big(\sum_i a_{ij}\Big)^2 \end{aligned}$$

is 0.

We have $|A^T x| \leqslant |x|$ for all x. With $x = (1, 1, \ldots, 1)^T$ this yields

$$\sum_j \Big(\sum_i a_{ij} \Big)^2 \leqslant |x|^2 = n,$$

from which it follows that $S \leqslant 0$. Hence $S = 0$.

Also solved by David K. Baxter, A. J. Bosch (The Netherlands), C. S. K. Chetty (India), Jesse Deutsch, H. Kestelman (England), Mauri Koskela (Finland), O. P. Lossers (The Netherlands), Marvin Marcus, A. McD. Mercer (Canada), Jorma Kaarlo Merikoski (Finland), Marcel F. Neuts, Paul J. Nikolai, Nicholas Passell, Mary Beth Ruskai, C. L. Thompson (England), John Tripp, Nam-Kiu Tsing (Hong Kong), Pei Yuan Wu (Taiwan), and the proposer.

The Error Term in Simpson's Rule

6368 [1981, 768]. *Proposed by Robert E. Shafer, Berkeley, Calif.*

Consider the integral

$$I = \int_{v/2}^{(v+1)/2} \psi(z)\, dz = \log\Big[\Gamma\Big(\frac{v+1}{2}\Big)/\Gamma(v/2)\Big], \mathrm{Re}\, v > 0$$

where ψ denotes the logarithmic derivative of the gamma function. Simpson's rule with two subintervals gives

$$I = \frac{1}{12}\psi\Big(\frac{v}{2}\Big) + \frac{1}{3}\psi\Big(\frac{v+1/2}{2}\Big) + \frac{1}{12}\psi\Big(\frac{v+1}{2}\Big) - E,$$

where

$$E = \frac{1}{92160}\psi^{(iv)}\Big(\frac{v}{2} + \frac{\theta}{2}\Big), 0 < \theta < 1.$$

Show that for real v, E is negative, by deriving the representation

$$E = \int_0^1 \left\{ \frac{1}{\log\Big(\frac{1+x}{1-x}\Big)} - \frac{1 + 2\sqrt{1-x^2}}{6x} \right\} \Big(\frac{1-x}{1+x}\Big)^{v-1} \frac{dx}{1+x}.$$

Solution by Lajos Takács, Case Western Reserve University. It is well known that, for $\mathrm{Re}\, z > 0$,

$$\log \Gamma(z) = \int_0^1 \Big(\frac{1 - u^{z-1}}{1-u} - z + 1 \Big) \frac{du}{\log u}$$

and

$$\psi(z) = \frac{d}{dz} \log \Gamma(z) = -\int_0^1 \Big(\frac{u^{z-1}}{1-u} + \frac{1}{\log u} \Big)\, du.$$

Substituting $u = y^2$ in these integrals, we can express

$$E = \frac{1}{12}\psi\Big(\frac{v}{2}\Big) + \frac{1}{3}\psi\Big(\frac{v}{2} + \frac{1}{4}\Big) + \frac{1}{12}\psi\Big(\frac{v}{2} + \frac{1}{2}\Big) - \log \frac{\Gamma\Big(\frac{v}{2} + \frac{1}{2}\Big)}{\Gamma\Big(\frac{v}{2}\Big)}$$

in the form

$$E = \frac{1}{6}\int_0^1 g\Big(\frac{1}{y}\Big) \frac{y^v}{(1-y^2)\log y}\, dy$$

for $\mathrm{Re}\, v > 0$, where

$$g(u) = (u + 4\sqrt{u} + 1)\log u - 6(u-1).$$

We have $|A^Tx| \leqslant |x|$ for all x. With $x = (1, 1, \ldots, 1)^T$ this yields

$$\sum_j \Big(\sum_i a_{ij} \Big)^2 \leqslant |x|^2 = n,$$

from which it follows that $S \leqslant 0$. Hence $S = 0$.

Also solved by David K. Baxter, A. J. Bosch (The Netherlands), C. S. K. Chetty (India), Jesse Deutsch, H. Kestelman (England), Mauri Koskela (Finland), O. P. Lossers (The Netherlands), Marvin Marcus, A. McD. Mercer (Canada), Jorma Kaarlo Merikoski (Finland), Marcel F. Neuts, Paul J. Nikolai, Nicholas Passell, Mary Beth Ruskai, C. L. Thompson (England), John Tripp, Nam-Kiu Tsing (Hong Kong), Pei Yuan Wu (Taiwan), and the proposer.

The Error Term in Simpson's Rule

6368 [1981, 768]. *Proposed by Robert E. Shafer, Berkeley, Calif.*

Consider the integral

$$I = \int_{v/2}^{(v+1)/2} \psi(z)\, dz = \log\Big[\Gamma\Big(\frac{v+1}{2}\Big) / \Gamma(v/2)\Big], \mathrm{Re}\, v > 0$$

where ψ denotes the logarithmic derivative of the gamma function. Simpson's rule with two subintervals gives

$$I = \frac{1}{12}\psi\Big(\frac{v}{2}\Big) + \frac{1}{3}\psi\Big(\frac{v+1/2}{2}\Big) + \frac{1}{12}\psi\Big(\frac{v+1}{2}\Big) - E,$$

where

$$E = \frac{1}{92160}\psi^{(iv)}\Big(\frac{v}{2} + \frac{\theta}{2}\Big), 0 < \theta < 1.$$

Show that for real v, E is negative, by deriving the representation

$$E = \int_0^1 \left\{ \frac{1}{\log\Big(\dfrac{1+x}{1-x}\Big)} - \frac{1 + 2\sqrt{1-x^2}}{6x} \right\} \Big(\frac{1-x}{1+x}\Big)^{v-1} \frac{dx}{1+x}.$$

Solution by Lajos Takács, Case Western Reserve University. It is well known that, for $\mathrm{Re}\, z > 0$,

$$\log \Gamma(z) = \int_0^1 \Big(\frac{1 - u^{z-1}}{1-u} - z + 1 \Big) \frac{du}{\log u}$$

and

$$\psi(z) = \frac{d}{dz} \log \Gamma(z) = -\int_0^1 \Big(\frac{u^{z-1}}{1-u} + \frac{1}{\log u} \Big)\, du.$$

Substituting $u = y^2$ in these integrals, we can express

$$E = \frac{1}{12}\psi\Big(\frac{v}{2}\Big) + \frac{1}{3}\psi\Big(\frac{v}{2} + \frac{1}{4}\Big) + \frac{1}{12}\psi\Big(\frac{v}{2} + \frac{1}{2}\Big) - \log \frac{\Gamma\Big(\dfrac{v}{2} + \dfrac{1}{2}\Big)}{\Gamma\Big(\dfrac{v}{2}\Big)}$$

in the form

$$E = \frac{1}{6} \int_0^1 g\Big(\frac{1}{y}\Big) \frac{y^v}{(1-y^2)\log y}\, dy$$

for $\mathrm{Re}\, v > 0$, where

$$g(u) = (u + 4\sqrt{u} + 1)\log u - 6(u-1).$$

Substituting $y = (1 - x)/(1 + x)$ in the above integral, we obtain the desired formula for E.
Next we observe that, for $t > 0$,

$$e^{-t}g(e^{2t}) = 4t\cosh t + 8t - 12\,\mathrm{sh}\,t$$
$$= 4\sum_{n=1}^{\infty}\frac{t^{2n+1}}{(2n)!}\left(1 - \frac{3}{2n+1}\right) > 0.$$

Hence $g(u) > 0$ for $u > 1$, and this implies that $E < 0$ when $v > 0$.

Also solved by Otto G. Ruehr and the proposer. M. L. Glasser and O. P. Lossers (The Netherlands) derived the desired representation for E but did not show that $E < 0$ for $v > 0$. M. E. Muldoon (Canada) observed that an explicit representation for E is not required to prove that $E < 0$ when $v > 0$, since this follows from the known formula

$$\psi^{(iv)}(x) = -24\sum_{n=0}^{\infty}(x+n)^{-5}, \quad x > 0.$$

The Sum $\Sigma_{k=1}^{N} k^{-1/2}\log k$

6371 [1981, 769]. *Proposed by Bruce C. Berndt, University of Illinois, Urbana-Champaign.*

Show that

$$\lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-1/2}\log k - 2\sqrt{N}\log N + 4\sqrt{N}\right) = -\zeta(1/2)\{\pi/4 + \gamma/2 + (\log(8\pi))/2\}$$

where $\zeta(s)$ denotes the Riemann zeta-function and γ denotes Euler's constant. Note

$$\zeta(1/2) = (2^{1/2}+1)\sum_{1}^{\infty}(-1)^{n}n^{-1/2}.$$

Solution by Otto G. Ruehr, Michigan Technological University, Houghton, Michigan. G. H. Hardy (*Divergent Series*, Oxford, 1949, pp. 333–334) employed the Euler-Maclaurin sum formula to show that

$$(*) \qquad \lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-s}\log k - \frac{N^{1-s}\log N}{1-s} + \frac{N^{1-s}}{(1-s)^2} - \frac{N^{-s}\log N}{2}\right) = -\zeta'(s)$$

for $\operatorname{Re} s > -1$. To calculate $\zeta'(1/2)$ we differentiate logarithmically the well-known identity

$$\zeta(s) = 2(2\pi)^{s-1}\sin\frac{\pi s}{2}\Gamma(1-s)\zeta(1-s)$$

and set $s = 1/2$, noting that $\Gamma'(1/2)/\Gamma(1/2) = -\gamma - \log 4$. The indicated result follows by taking $s = 1/2$ in (*).

Also solved by Paul S. Bruckman, P. J. Forrester (Australia), M. L. Glasser, Gaston H. Gonnet (Canada), Nathaniel Grossman, Sidney Heller, H. Jager (The Netherlands), O. P. Lossers (The Netherlands), William A. Newcomb, Helmut Prodinger (Austria), Don Redmond, Hermann Schmidt (West Germany), Robert E. Shafer, David Shelupsky, Lajos Takács, L. Van Hamme (Belgium), and the proposer.

An L_p-Norm Inequality

6372 [1981, 769]. *Proposed by David R. Brillinger, University of California, Berkeley.*

Let n be a positive integer and f in $L_p(R)$ with $p = (n+1)/n$. Then

$$\int\cdots\int|f(x_1)\cdots f(x_n)f(x_1+\cdots+x_n)|dx_1\ldots dx_n \leqslant \|f\|_p^{n+1}.$$

Substituting $y = (1 - x)/(1 + x)$ in the above integral, we obtain the desired formula for E.

Next we observe that, for $t > 0$,

$$e^{-t}g(e^{2t}) = 4t\cosh t + 8t - 12\,\mathrm{sh}\,t$$
$$= 4\sum_{n=1}^{\infty}\frac{t^{2n+1}}{(2n)!}\left(1 - \frac{3}{2n+1}\right) > 0.$$

Hence $g(u) > 0$ for $u > 1$, and this implies that $E < 0$ when $v > 0$.

Also solved by Otto G. Ruehr and the proposer. M. L. Glasser and O. P. Lossers (The Netherlands) derived the desired representation for E but did not show that $E < 0$ for $v > 0$. M. E. Muldoon (Canada) observed that an explicit representation for E is not required to prove that $E < 0$ when $v > 0$, since this follows from the known formula

$$\psi^{(iv)}(x) = -24\sum_{n=0}^{\infty}(x+n)^{-5}, \quad x > 0.$$

The Sum $\sum_{k=1}^{N} k^{-1/2}\log k$

6371 [1981, 769]. *Proposed by Bruce C. Berndt, University of Illinois, Urbana-Champaign.*

Show that

$$\lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-1/2}\log k - 2\sqrt{N}\log N + 4\sqrt{N}\right) = -\zeta(1/2)\{\pi/4 + \gamma/2 + (\log(8\pi))/2\}$$

where $\zeta(s)$ denotes the Riemann zeta-function and γ denotes Euler's constant. Note

$$\zeta(1/2) = (2^{1/2} + 1)\sum_{1}^{\infty}(-1)^n n^{-1/2}.$$

Solution by Otto G. Ruehr, Michigan Technological University, Houghton, Michigan. G. H. Hardy (*Divergent Series*, Oxford, 1949, pp. 333–334) employed the Euler-Maclaurin sum formula to show that

$$(*)\qquad \lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-s}\log k - \frac{N^{1-s}\log N}{1-s} + \frac{N^{1-s}}{(1-s)^2} - \frac{N^{-s}\log N}{2}\right) = -\zeta'(s)$$

for Re $s > -1$. To calculate $\zeta'(1/2)$ we differentiate logarithmically the well-known identity

$$\zeta(s) = 2(2\pi)^{s-1}\sin\frac{\pi s}{2}\Gamma(1-s)\zeta(1-s)$$

and set $s = 1/2$, noting that $\Gamma'(1/2)/\Gamma(1/2) = -\gamma - \log 4$. The indicated result follows by taking $s = 1/2$ in (*).

Also solved by Paul S. Bruckman, P. J. Forrester (Australia), M. L. Glasser, Gaston H. Gonnet (Canada), Nathaniel Grossman, Sidney Heller, H. Jager (The Netherlands), O. P. Lossers (The Netherlands), William A. Newcomb, Helmut Prodinger (Austria), Don Redmond, Hermann Schmidt (West Germany), Robert E. Shafer, David Shelupsky, Lajos Takács, L. Van Hamme (Belgium), and the proposer.

An L_p-Norm Inequality

6372 [1981, 769]. *Proposed by David R. Brillinger, University of California, Berkeley.*

Let n be a positive integer and f in $L_p(R)$ with $p = (n+1)/n$. Then

$$\int\cdots\int|f(x_1)\cdots f(x_n)f(x_1+\cdots+x_n)|dx_1\ldots dx_n \leqslant \|f\|_p^{n+1}.$$

Substituting $y = (1 - x)/(1 + x)$ in the above integral, we obtain the desired formula for E.

Next we observe that, for $t > 0$,

$$e^{-t}g(e^{2t}) = 4t\cosh t + 8t - 12\,\mathrm{sh}\,t$$
$$= 4\sum_{n=1}^{\infty}\frac{t^{2n+1}}{(2n)!}\left(1 - \frac{3}{2n+1}\right) > 0.$$

Hence $g(u) > 0$ for $u > 1$, and this implies that $E < 0$ when $v > 0$.

Also solved by Otto G. Ruehr and the proposer. M. L. Glasser and O. P. Lossers (The Netherlands) derived the desired representation for E but did not show that $E < 0$ for $v > 0$. M. E. Muldoon (Canada) observed that an explicit representation for E is not required to prove that $E < 0$ when $v > 0$, since this follows from the known formula

$$\psi^{(iv)}(x) = -24\sum_{n=0}^{\infty}(x+n)^{-5}, \quad x > 0.$$

The Sum $\sum_{k=1}^{N} k^{-1/2}\log k$

6371 [1981, 769]. *Proposed by Bruce C. Berndt, University of Illinois, Urbana-Champaign.*

Show that

$$\lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-1/2}\log k - 2\sqrt{N}\log N + 4\sqrt{N}\right) = -\zeta(1/2)\{\pi/4 + \gamma/2 + (\log(8\pi))/2\}$$

where $\zeta(s)$ denotes the Riemann zeta-function and γ denotes Euler's constant. Note

$$\zeta(1/2) = (2^{1/2} + 1)\sum_{1}^{\infty}(-1)^n n^{-1/2}.$$

Solution by Otto G. Ruehr, Michigan Technological University, Houghton, Michigan. G. H. Hardy (*Divergent Series*, Oxford, 1949, pp. 333–334) employed the Euler-Maclaurin sum formula to show that

$$(*)\qquad \lim_{N\to\infty}\left(\sum_{k=1}^{N} k^{-s}\log k - \frac{N^{1-s}\log N}{1-s} + \frac{N^{1-s}}{(1-s)^2} - \frac{N^{-s}\log N}{2}\right) = -\zeta'(s)$$

for $\mathrm{Re}\, s > -1$. To calculate $\zeta'(1/2)$ we differentiate logarithmically the well-known identity

$$\zeta(s) = 2(2\pi)^{s-1}\sin\frac{\pi s}{2}\Gamma(1-s)\zeta(1-s)$$

and set $s = 1/2$, noting that $\Gamma'(1/2)/\Gamma(1/2) = -\gamma - \log 4$. The indicated result follows by taking $s = 1/2$ in (*).

Also solved by Paul S. Bruckman, P. J. Forrester (Australia), M. L. Glasser, Gaston H. Gonnet (Canada), Nathaniel Grossman, Sidney Heller, H. Jager (The Netherlands), O. P. Lossers (The Netherlands), William A. Newcomb, Helmut Prodinger (Austria), Don Redmond, Hermann Schmidt (West Germany), Robert E. Shafer, David Shelupsky, Lajos Takács, L. Van Hamme (Belgium), and the proposer.

An L_p-Norm Inequality

6372 [1981, 769]. *Proposed by David R. Brillinger, University of California, Berkeley.*

Let n be a positive integer and f in $L_p(R)$ with $p = (n+1)/n$. Then

$$\int\cdots\int|f(x_1)\cdots f(x_n)f(x_1+\cdots+x_n)|dx_1\ldots dx_n \leqslant \|f\|_p^{n+1}.$$

(This inequality yields directly the result of Problem 5314, this MONTHLY, 72 (1965) 795, discussed by H. Dym, this MONTHLY, 87 (1980) 53-54).

Solution by Pei Yuan Wu, Department of Applied Mathematics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China. By Hölder's inequality, the integral in question does not exceed the following product of $n + 1$ integrals

$$\left(\int \cdots \int |f(x_2) \cdots f(x_n) f(x_1 + \cdots + x_n)|^p dx_1 \ldots dx_n\right)^{1/(n+1)} \cdots$$
$$\left(\int \cdots \int |f(x_1) \cdots f(x_{n-1}) f(x_1 + \cdots + x_n)|^p dx_1 \cdots dx_n\right)^{1/(n+1)}$$
$$\left(\int \cdots \int |f(x_1) \cdots f(x_n)|^p dx_1 \ldots dx_n\right)^{1/(n+1)}$$
$$= \left(\int |f(x)|^p dx\right)^{n^2/(n+1)} \left(\int |f(x)|^p dx\right)^{n/(n+1)} \quad \text{(by Fubini's theorem)}$$
$$= \|f\|_p^{n+1}.$$

Also solved by Ulrich Abel (West Germany), K. F. Andersen (Canada), G. A. Edgar, Noel Glick, O. P. Lossers (The Netherlands), Nicholas Miller, Michael D. Perlman, Jay Warendorff, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Turtle Geometry. The Computer as a Medium for Exploring Mathematics. By Harold Abelson and Andrea diSessa. The M.I.T. Press, Cambridge, MA, 1981. xx + 477 pp., $20.00.

GEORGE K. FRANCIS
Department of Mathematics, University of Illinois, Urbana, IL 61801

Astronomer, surveyor, mason, and carpenter measure heaven and earth, and create a world of parallels and perpendiculars. One imagines Plato handing his students a piece of string and instructing them to go and discover the rules of geometry on the dusty grounds of his academy. Straight lines connect here to there, and great circular arcs trace across the imagination of pupils for 2,000 years as they learn their Euclid. Instruments of measurement improve. Astrolabe and telescope chart the cosmos, drafting tools and paper aid artist and architect to explore perspective. Euclid yields to Descartes and Riemann. Geometry acquires hyphens: analytic-, projective-, algebraic-, differential-, and gives birth to topology. Once called the analysis of place, this offspring soon escapes its narrow mission and invades the furthest corners of science, from logic (ultra filters) to linguistics (catastrophe theory).

Yet, a century after the great synthesis of analysis, geometry and algebra that culminated in the work of Klein and Poincaré, the exciting ideas of Thurston *et al.* have inaugurated a rennaissance of interest in low-dimensional geometry. In order to appreciate and understand the Thurston program, and perhaps contribute to it, a profound familiarity with the most elementary concepts, such as curvature, connectivity, and groups of structure-preserving transformations, is essential. But how can we provide a cohort of native speakers of a geometrical language which is currently learned only in graduate school?

(This inequality yields directly the result of Problem 5314, this MONTHLY, 72 (1965) 795, discussed by H. Dym, this MONTHLY, 87 (1980) 53-54).

Solution by Pei Yuan Wu, Department of Applied Mathematics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China. By Hölder's inequality, the integral in question does not exceed the following product of $n+1$ integrals

$$\left(\int\cdots\int|f(x_2)\cdots f(x_n)f(x_1+\cdots+x_n)|^p dx_1\ldots dx_n\right)^{1/(n+1)}\cdots$$
$$\left(\int\cdots\int|f(x_1)\cdots f(x_{n-1})f(x_1+\cdots+x_n)|^p dx_1\cdots dx_n\right)^{1/(n+1)}$$
$$\left(\int\cdots\int|f(x_1)\cdots f(x_n)|^p dx_1\ldots dx_n\right)^{1/(n+1)}$$
$$=\left(\int|f(x)|^p dx\right)^{n^2/(n+1)}\left(\int|f(x)|^p dx\right)^{n/(n+1)}\quad\text{(by Fubini's theorem)}$$
$$=\|f\|_p^{n+1}.$$

Also solved by Ulrich Abel (West Germany), K. F. Andersen (Canada), G. A. Edgar, Noel Glick, O. P. Lossers (The Netherlands), Nicholas Miller, Michael D. Perlman, Jay Warendorff, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Turtle Geometry. The Computer as a Medium for Exploring Mathematics. By Harold Abelson and Andrea diSessa. The M.I.T. Press, Cambridge, MA, 1981. xx + 477 pp., $20.00.

GEORGE K. FRANCIS
Department of Mathematics, University of Illinois, Urbana, IL 61801

Astronomer, surveyor, mason, and carpenter measure heaven and earth, and create a world of parallels and perpendiculars. One imagines Plato handing his students a piece of string and instructing them to go and discover the rules of geometry on the dusty grounds of his academy. Straight lines connect here to there, and great circular arcs trace across the imagination of pupils for 2,000 years as they learn their Euclid. Instruments of measurement improve. Astrolabe and telescope chart the cosmos, drafting tools and paper aid artist and architect to explore perspective. Euclid yields to Descartes and Riemann. Geometry acquires hyphens: analytic-, projective-, algebraic-, differential-, and gives birth to topology. Once called the analysis of place, this offspring soon escapes its narrow mission and invades the furthest corners of science, from logic (ultra filters) to linguistics (catastrophe theory).

Yet, a century after the great synthesis of analysis, geometry and algebra that culminated in the work of Klein and Poincaré, the exciting ideas of Thurston *et al.* have inaugurated a rennaissance of interest in low-dimensional geometry. In order to appreciate and understand the Thurston program, and perhaps contribute to it, a profound familiarity with the most elementary concepts, such as curvature, connectivity, and groups of structure-preserving transformations, is essential. But how can we provide a cohort of native speakers of a geometrical language which is currently learned only in graduate school?

Abelson and diSessa's "Turtle Geometry" represents a pedagogical program that is a vital first step in this direction. In a word, it says: computergraphics. There is no question that this truly new tool of visualization and its associated syntax (procedure oriented recursive programming) could be a direct access to an intuitive understanding of matters generally considered very difficult. Just how this might be done is expertly explored in this primer of computer based geometry.

Undoubtedly, the authors mean to take their readers by the hand and guide them through a wonderful safari to the curved space of General Relativity. But this book is no "Mr. Thompson in Wonderland." There are eight chapters of sound training in plane, space, and surface geometry preceding it. Their turtle, named for Walter's robot tortoises that were programmed to advance and turn like the Milton-Bradley programmable toy "Big Trak," is secretly a point (= state) in the principal $SO(n)$ bundle over a manifold. In the plane (Chapters 1, 2, 4) the turtle has a direction as well as a position. It changes these by obeying a string made up of the two atomic commands: Fs = "move s units from the current position in the direction of the current heading" and $R\theta$ = "rotate the heading θ angular units from the current direction." Thus, over a turtle path, ("tour" would have been better because its program is more of an itinerary than a highway), it accumulates total turning as well as total displacement.

To give a sample of turtle pedagogy, let me trace their "poly closing theorem" through the text and certain exercises, which they provide with hints and an explanation of their solution. The theorem claims that simple iteration of $FsR\theta$ returns the turtle to its initial state if and when the total turning reaches an integral multiple of 2π. Following ample experiments in support of its truth, a first proof is sketched. It depends on the observation that the vertices of the polygonal trail left by this tour lie on a common circle of reciprocal radius equal to $|2\sin(\theta/2)|/s$. The authors resist the temptation of proving this (elementary) lemma, or of calling this ratio the curvature of the circle, for a very good reason. In orthodox turtle thought, based on the radical subjectivity of strictly local (intrinsic) experiences, a turtle circle is not so much the locus of positions equidistant from a point, but the (ideal) path approximated by tours with decreasing parameters but constant ratio θ/s. Though called "curvature" early in the book, the authors avoid this term in relation to curves because they do not intend to define surface curvature in terms of the Gauss map [**5**]. Just how they manage to sidestep the *Theorema Egregium* and motivate intrinsic curvature from the start is one of the more ingenious if controversial features of Turtle Geometry.

The second proof argues (by contradiction) that if after n steps, headings agree (as they must for θ a proper rational fraction of 2π) but at different positions, $P0 \neq Pn = (FsR\theta)^n(P0)$, then the turtle must wander off to infinity. It does so inside a strip along the line $P0Pn$, which it meets every n steps, never straying farther than $ns/2$ from it in between. On the other hand, applying the argument to the initial state $(P1, H1 = H0 + \theta)$ instead, obliges the turtle to wander off along another road rotated from the former, which no plane turtle can do.

After two chapters of thorough motivation, vector geometry is presented as the natural algebraic tool for implementing turtle programs on real computers. Now the classic vector proof, using isotropy and linearity of rotations can be explored. But real geometric insight is generated in an exercise (with hint and answer) concerning the failure of this theorem in three space. Here an aquatic turtle, or perhaps a bird, has an orthogonal frame for its heading, and rotations have three primitives appropriately called pitch, roll and yaw. The second argument here leads not to a contradiction but the existence of invariant axes for elements of SO(3). The polybird happily spirals off to infinity along a helix.

The central character of Turtle Geometry, however, is the "simple closed path theorem" to the effect that whatever complicated program generates a turtle loop that never crosses itself, the total turning is $+2\pi$; compare Hopf [**6**]. Its demonstrable failure on a sphere leads to the notion of angular excess and eventually to intrinsic curvature density. The local form of the Gauss-Bonnet theorem [**3**, p. 268] serves as a quasi definition of the total curvature of a simply connected surface patch. Labor is rewarded with a thoroughly convincing demonstration of the Kronecker-Dyck

theorem: for a closed surface the total curvature is 2π times the Euler characteristic [7], [4].

One may well wonder how all this can be done without reference to the calculus and with any semblance of precision. On the matter of rigor, always a touchy point with "elementary" books, let me suggest a humanistic criterion. An explanation is a discourse that satisfies the curiosity of the hearer and a proof is an argument that removes the obstacles to belief. Those definitions and proofs that fit the level of combinatorial sophistication expected of the reader are as sound as a fiddle. Among these is an algorithm teaching the turtle to escape a maze based on Hopf's theorem. Another is a proof of Hopf's theorem for a noncrossing right angled turtle loop by collapsing it to a square. Other, more heuristic explanations arouse an appetite and prepare the way for the analysis and topology needed to cement the gaps. One reason the authors can maintain plausibility lies in the digitized nature of turtle paths. Since the computer turtle always has some heading, the ideal curve "pointed to" by the traces on the (flat!) computer screen is piecewise regular (= nonzero velocity) in the sense of Hopf [6] and Whitney [12]. A broken geodesic on a curved or polyhedral surface becomes an "even strided turtle walk," or better, a wheeled turtle that turns its right and left wheels at the same speed on the forward motion. Parallel translation of a fixed direction relative to its initial heading (the affine connection) is managed by making the appropriate additive corrections at the turns. Transition between adjacent charts (... what to do when the turtle reaches the screen edge) in an atlas for a manifold is nicely motivated, for example, by taking the turtle for a walk on a cube.

And so on. I will not spoil the fun of the chase by revealing all the geometrical concepts amenable to turtle heuristics. The book is magnificently illustrated. A plenitude of problems is helped by separate appendices for hints and answers. Turtle Procedure Language, the pseudocode used throughout the text, is fairly clear and an appendix ties it to the common speech of APPLEBASIC and PASCAL as well as its parent LOGO. The only serious flaw is the absence of a bibliography. To the undergraduate veteran of Turtle Geometry I recommend doCarmo's superb text [3]. For the geometrical convert to turtle pedagogy there are several good projects left virtually untouched by this volume. One that springs to mind is to explore also the extrinsic curvature of surfaces in 3-space, along the lines of Tom Banchoff's films [1].

In closing, I should like to spin a yarn from Gauss to Nelson Max's computer graphics masterpiece on the sphere eversion [8]. One thread of it is taken up in Chapter 4 on deforming plane closed turtle paths. The rest could lead to a computer based geometry whose protagonist is an entire surface writhing in higher dimensions.

A succession of small changes in the parameters of a turtle loop simulates a regular homotopy, provided that only angles $|\theta| < \pi$ are used in programming the turtle. The resulting deformation preserves the total turning of the curve. Graustein's elegant analytic proof in [12], that (conversely) two regular closed plane curves with equal total turning are regularly homotopic, was generalized by Smale. In his thesis [10] he identifies the fundamental group, $\pi_1(TS(M^n), *)$, of the direction (= unit sphere) bundle over a Riemannian manifold as the set $\pi_0(\mathrm{Im}(S^1, M; *))$ of regular homotopy classes of immersions of the circle in M with common initial place and heading. For $n = 2$ and M the plane, $TS(R^2) = R^2 \times S^1$ and the invariant is the total turning divided by 2π. (This number, called "amplitudo" by Gauss, has acquired many aliases since.) In a sequel [11], Smale shows that the regular homotopy classes, $\pi_0(\mathrm{Im}(S^2, R^n))$, of immersions of the sphere are in $1:1$ correspondence with $\pi_2(V_{n,2})$, where $V_{n,k}$ is the Stiefel manifold of k-frames in n-space. Since the cross product of two independent vectors in R^3 determines an independent third, and the triplet can be orthonormalized, $V_{3,2}$ retracts to SO(3). Since S^3 doubly covers SO(3) and $\pi_2(S^3)$ is trivial, all immersions of the sphere in space are connected to each other. A regular homotopy from the standard inclusion to its central (= antipodal) reflection is called an eversion of the sphere.

The notion of regular homotopy and the Whitney-Graustein theorem have little known precursors in the 1901 dissertation of Hilbert's student Werner Boy [2]. His curves in the plane and surfaces in space are still the 19th century loci of algebraic constraints, not the images of

mappings. His "completely continuous deformations," restricted only by an upper bound on the absolute value of the Gaussian curvature measure, are not always (equivalent to) a regular homotopy, though for circles and spheres it is. Boy's highly visual proof that total curvature classifies the deformations of plane closed loops, which could have a pretty turtle version, is omitted from his published paper, with good reason. Generally, its surface analog is false. A strip of paper joined after one full twist (not a Möbius band) will not deform into an untwisted one, though both have zero curvature. (In view of Smale's theorem, one must seek a counterexample for closed surfaces among more complicated ones than spheres.) Boy proceeds with a geometrical realization of Gauss' plan of relating topological connectivity to total curvature. He develops an equivalent to curvature that works also for piecewise polyhedral surfaces, reminiscent of Turtle Geometry, Chapters 5–8. He constructs his celebrated immersion of the projective plane as the missing link ($\chi = 1$, $n = 0$) in the zoo of known surfaces in real 3-space for which

$$C = 2\pi\chi + 4\pi n$$

where C is the total (polyhedral) curvature, χ the Euler-Poincaré characteristic and n the number of double curves ending in pinchpoints. This beautiful surface has been with us for 80 years, and only recently, Petit and Souriau [**9**], working out an idea of Bernard Morin, found a parametrization that works well on an APPLE II.

References

1. T. Banchoff, The Gaussmap, Banchoff-Strauss Productions, Providence, RI, 1977.

2. W. Boy, Ueber die Curvatura integra u.d. Topologie geschlossener Flächen, dissertation (Hilbert), Göttingen, 1901; summarized in Math. Ann., 57(1903).

3. M. doCarmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.

4. W. Dyck, Analysis situs l, Math. Ann., 32(1888).

5. K. F. Gauss, Disquitiones generales circa superficies curvas, Comm. Soc. Göttingen, 6(1823); translated in General Investigations of Curved Surfaces, Raven Press, NY, 1965.

6. H. Hopf, Ueber die Drehung der Tangenten und Sehnen ebener Kurven, Compositio Math., 2(1935).

7. L. Kronecker, Ueber Systeme von Funktionen mehrerer Variablen, Berliner Monatsbericht, Aug. 1869.

8. N. Max, Turning the Sphere Inside Out, International Film Bureau, Chicago 1977.

9. J. Petit and J. Souriau, La Surface de Boy, La Recherche, 132 Avril 1982; abstract in C. R. Acad. Sci. Paris, 293(1981).

10. S. Smale, Regular curves on Riemannian manifolds, dissertation (Bott), Trans. Amer. Math. Soc., 81(1958).

11. S. Smale, A classification of immersions of the two-sphere, Trans. Amer. Math. Soc., 90(1959).

12. H. Whitney, On regular closed curves in the plane, Compositio Math., 4(1937).

Mathematics and Physics. By Yu. I. Manin. Translated by Ann and Neal Koblitz. Birkhauser, Boston, 1981. xii + 99 pp., $10.00.

R. O. WELLS, JR.

Department of Mathematics, Rice University, Houston, TX 77251

In the past 10 years there has been a noticeable increase in the interaction between mathematicians and physicists which seems reminiscent of the profound interaction which took place in the developments in quantum mechanics in the first third of this century. Today we have theoretical physicists learning about Chern classes, Pontryagin numbers, $K3$-surfaces, projective algebraic geometry, while mathematicians trained in areas of number theory, algebraic geometry, or several complex variables are involved in learning about aspects of relativity theory or quantum field theory, and the modern theory of elementary particles. From the mathematician's point of view, the areas of modern mathematics which seem to be having a new impact on modern physics include: differential geometry, algebraic geometry, algebraic topology, several complex variables,

mappings. His "completely continuous deformations," restricted only by an upper bound on the absolute value of the Gaussian curvature measure, are not always (equivalent to) a regular homotopy, though for circles and spheres it is. Boy's highly visual proof that total curvature classifies the deformations of plane closed loops, which could have a pretty turtle version, is omitted from his published paper, with good reason. Generally, its surface analog is false. A strip of paper joined after one full twist (not a Möbius band) will not deform into an untwisted one, though both have zero curvature. (In view of Smale's theorem, one must seek a counterexample for closed surfaces among more complicated ones than spheres.) Boy proceeds with a geometrical realization of Gauss' plan of relating topological connectivity to total curvature. He develops an equivalent to curvature that works also for piecewise polyhedral surfaces, reminiscent of Turtle Geometry, Chapters 5–8. He constructs his celebrated immersion of the projective plane as the missing link ($\chi = 1$, $n = 0$) in the zoo of known surfaces in real 3-space for which

$$C = 2\pi\chi + 4\pi n$$

where C is the total (polyhedral) curvature, χ the Euler-Poincaré characteristic and n the number of double curves ending in pinchpoints. This beautiful surface has been with us for 80 years, and only recently, Petit and Souriau [**9**], working out an idea of Bernard Morin, found a parametrization that works well on an APPLE II.

References

1. T. Banchoff, The Gaussmap, Banchoff-Strauss Productions, Providence, RI, 1977.

2. W. Boy, Ueber die Curvatura integra u.d. Topologie geschlossener Flächen, dissertation (Hilbert), Göttingen, 1901; summarized in Math. Ann., 57(1903).

3. M. doCarmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.

4. W. Dyck, Analysis situs I, Math. Ann., 32(1888).

5. K. F. Gauss, Disquitiones generales circa superficies curvas, Comm. Soc. Göttingen, 6(1823); translated in General Investigations of Curved Surfaces, Raven Press, NY, 1965.

6. H. Hopf, Ueber die Drehung der Tangenten und Sehnen ebener Kurven, Compositio Math., 2(1935).

7. L. Kronecker, Ueber Systeme von Funktionen mehrerer Variablen, Berliner Monatsbericht, Aug. 1869.

8. N. Max, Turning the Sphere Inside Out, International Film Bureau, Chicago 1977.

9. J. Petit and J. Souriau, La Surface de Boy, La Recherche, 132 Avril 1982; abstract in C. R. Acad. Sci. Paris, 293(1981).

10. S. Smale, Regular curves on Riemannian manifolds, dissertation (Bott), Trans. Amer. Math. Soc., 81(1958).

11. S. Smale, A classification of immersions of the two-sphere, Trans. Amer. Math. Soc., 90(1959).

12. H. Whitney, On regular closed curves in the plane, Compositio Math., 4(1937).

Mathematics and Physics. By Yu. I. Manin. Translated by Ann and Neal Koblitz. Birkhauser, Boston, 1981. xii + 99 pp., $10.00.

R. O. WELLS, JR.

Department of Mathematics, Rice University, Houston, TX 77251

In the past 10 years there has been a noticeable increase in the interaction between mathematicians and physicists which seems reminiscent of the profound interaction which took place in the developments in quantum mechanics in the first third of this century. Today we have theoretical physicists learning about Chern classes, Pontryagin numbers, $K3$-surfaces, projective algebraic geometry, while mathematicians trained in areas of number theory, algebraic geometry, or several complex variables are involved in learning about aspects of relativity theory or quantum field theory, and the modern theory of elementary particles. From the mathematician's point of view, the areas of modern mathematics which seem to be having a new impact on modern physics include: differential geometry, algebraic geometry, algebraic topology, several complex variables,

and others. This list doesn't include the traditional interaction with such fields as operator algebras, group representation theory, and partial differential equations, etc. Contemporary mathematicians who have become involved in this new interaction include, for instance, M. F. Atiyah, S. Novikov, and I. Singer, as well as the author of the book being reviewed. Physicists involved in the interaction include, for instance, G. t'Hooft, (elementary particle theory), and R. Penrose (relativity theory), as well as many others.

The interaction between mathematics and physics goes back to the beginnings of both intellectual disciplines, and the cross-fertilization has had enormous benefits for both areas. However, they are indeed distinct disciplines with their own standards, methodology, and fundamental problems. The book under review is written by a mathematician who had been asked by a certain forum to talk about the interaction between these two fields, and this published book is an edited version of a series of lectures on the subject. The result is a delightful experience written in a beautiful style which is not at all pedantic, and which attempts to indicate the diversity of the topic as well as the depth and nature of the interaction. In the first paragraph of the book the author relates an anecdote to the effect that a well-known professor of logic was in the habit of beginning a course in logic by asserting that "logic is the science of laws of thought; now I must tell you what the words 'science', 'laws', and 'thought' mean, but I won't tell you what 'of' means." Manin, in his book, also doesn't attempt to explain the word "and" in the title of his book. He discusses, in a stream of consciousness fashion, numerous topics from both mathematics and physics and the reader can divine for himself the nature of the word "and." The table of contents illustrates the breadth of the book: (1) a bird's eye view of mathematics; (2) physical quantities, dimensions and constants: the source of numbers in physics; (3) a drop of milk: observer, observation, observable, and unobservable; (4) space-time as a physics system; (5) action and symmetry. In Chapter 1 he talks about set theory, mathematical truth, linearity, curvature, formulas and other mathematical notions. In Chapter 2 he discusses in a beautiful manner and in a language that mathematicians can appreciate the nature of physical laws and the importance of scales and fundamental units. In the next two chapters he takes an interesting excursion through the areas of quantum theory and relativity theory, and their interaction. He concludes with a brief discussion of symmetry and ends up with a tantalizing comparison between the theorem of Euclid on the infinitude of primes and the theorem of Deligne which solves an old, easily-stated conjecture of Ramanujan concerning prime numbers.

The book is beautifully written and will at times tax the reader whose knowledge doesn't reach far enough in a given direction, but the flavor of a vast variety of topics and their significance in both mathematics and physics is presented with the skill of a scientist who knows what it means to communicate with an audience. The one major defect in the book is the complete lack of references (such is the nature of a lecture), but the stimulus for further reading will certainly be there after reading the book. It is too bad that the author didn't attempt to bridge this additional gap of communication, as it would have added greatly to the book. Nevertheless, I recommend the book highly to the mathematics community (and the physics community as well!) as a significant essay on the nature of human knowledge in two areas which have a profound effect on the civilization we live in.

MISCELLANEA

107. Anyone who cannot cope with mathematics is not fully human.

—Robert A. Heinlein, *Time Enough for Love*, Berkley Medallion Edition, p. 247.

and others. This list doesn't include the traditional interaction with such fields as operator algebras, group representation theory, and partial differential equations, etc. Contemporary mathematicians who have become involved in this new interaction include, for instance, M. F. Atiyah, S. Novikov, and I. Singer, as well as the author of the book being reviewed. Physicists involved in the interaction include, for instance, G. t'Hooft, (elementary particle theory), and R. Penrose (relativity theory), as well as many others.

The interaction between mathematics and physics goes back to the beginnings of both intellectual disciplines, and the cross-fertilization has had enormous benefits for both areas. However, they are indeed distinct disciplines with their own standards, methodology, and fundamental problems. The book under review is written by a mathematician who had been asked by a certain forum to talk about the interaction between these two fields, and this published book is an edited version of a series of lectures on the subject. The result is a delightful experience written in a beautiful style which is not at all pedantic, and which attempts to indicate the diversity of the topic as well as the depth and nature of the interaction. In the first paragraph of the book the author relates an anecdote to the effect that a well-known professor of logic was in the habit of beginning a course in logic by asserting that "logic is the science of laws of thought; now I must tell you what the words 'science', 'laws', and 'thought' mean, but I won't tell you what 'of' means." Manin, in his book, also doesn't attempt to explain the word "and" in the title of his book. He discusses, in a stream of consciousness fashion, numerous topics from both mathematics and physics and the reader can divine for himself the nature of the word "and." The table of contents illustrates the breadth of the book: (1) a bird's eye view of mathematics; (2) physical quantities, dimensions and constants: the source of numbers in physics; (3) a drop of milk: observer, observation, observable, and unobservable; (4) space-time as a physics system; (5) action and symmetry. In Chapter 1 he talks about set theory, mathematical truth, linearity, curvature, formulas and other mathematical notions. In Chapter 2 he discusses in a beautiful manner and in a language that mathematicians can appreciate the nature of physical laws and the importance of scales and fundamental units. In the next two chapters he takes an interesting excursion through the areas of quantum theory and relativity theory, and their interaction. He concludes with a brief discussion of symmetry and ends up with a tantalizing comparison between the theorem of Euclid on the infinitude of primes and the theorem of Deligne which solves an old, easily-stated conjecture of Ramanujan concerning prime numbers.

The book is beautifully written and will at times tax the reader whose knowledge doesn't reach far enough in a given direction, but the flavor of a vast variety of topics and their significance in both mathematics and physics is presented with the skill of a scientist who knows what it means to communicate with an audience. The one major defect in the book is the complete lack of references (such is the nature of a lecture), but the stimulus for further reading will certainly be there after reading the book. It is too bad that the author didn't attempt to bridge this additional gap of communication, as it would have added greatly to the book. Nevertheless, I recommend the book highly to the mathematics community (and the physics community as well!) as a significant essay on the nature of human knowledge in two areas which have a profound effect on the civilization we live in.

MISCELLANEA

107. Anyone who cannot cope with mathematics is not fully human.

—Robert A. Heinlein, *Time Enough for Love*, Berkley Medallion Edition, p. 247.

LETTERS TO THE EDITOR

Material for this department should be sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

The "Atom Bomb" game (MISCELLANEA, November 1982, 687) was misquoted. In the correct version (which has no solution), each player picks a number in $[0, 1]$ and the one with the greater number kills the other, or in a tie both players survive—unless the sum is greater than *or equal to* 1, in which case the referee kills both. N. J. Fine and I heard this game at the Operations Evaluation Group (U. S. Navy) shortly after joining it in late 1945; it is possible that Erdős or Kaplansky heard it from one of us.

Here is a more realistic model I use in lectures, especially as a closing line. Since it deals with positive integers rather than the unit interval, general audiences are more comfortable (?) with it. Each player picks a positive integer and the player with the greater one kills the other, or in a tie both players survive—all this subject to the following condition: there is a number N, whose value is known to the referee *but is not known to either of the players*; if the sum is greater than N, then the referee kills them both.

Leonard Gillman
Department of Mathematics
University of Texas at Austin
Austin, TX 78712

Editor:

Does anybody know of any studies which show that a person with a degree in mathematics education makes a better high school teacher than a person with a straight degree in mathematics with no courses in education? I would also like to pose the same question with regard to chemistry, physics, biology, English, etc.

Since most people with a degree in mathematics education have fewer hours of mathematics than a person with a straight mathematics degree, it seems that the person with a straight mathematics degree would make a better high school teacher on the average, and particularly so from Algebra II onward. It also seems that the substitution of mathematics courses taught in the mathematics department in place of education and mathematics education courses taught elsewhere would yield a better high school teacher on the average.

Does anybody have any answers, ideas or comments?

Charles T. Scarborough
Department of Mathematics
and Statistics
Mississippi State University
Mississippi State, MS 39762

SOME INFORMATIVE AND USEFUL BOOKS FROM THE MAA...

Browse through this list and see if your library is missing some of these important books published by the Association.

APPLICATIONS OF UNDERGRADUATE MATHEMATICS IN ENGINEERING—written and edited by Ben Noble. A collection of articles prepared by engineers for the Committee on the Undergraduate Program in Mathematics. 364 pp. Hardbound.
List: $19.00. MAA Member: $13.00.

ANNOTATED BIBLIOGRAPHY OF EXPOSITORY WRITING IN THE MATHEMATICAL SCIENCES, prepared by M. P. Gaffney and L. A. Steen. An invaluable reference source of expository articles in mathematics. 282 pp. Paperbound.
List: $12.00. MAA Member: $8.00.

AN ANNOTATED BIBLIOGRAPHY OF FILMS AND VIDEOTAPES FOR COLLEGE MATHEMATICS, by David I. Schneider. An up-to-date listing of films and videotapes available for classroom use. 107 pp. Paperbound.
List: $9.00. MAA Members: $6.00.

A BASIC LIBRARY LIST FOR TWO-YEAR COLLEGES, prepared by the Committee on Basic Library Lists. A recommended library nucleus for two-year colleges. 66. pp. Paperbound.
List: $8.00. MAA Member: $6.00.

A BASIC LIBRARY LIST FOR FOUR-YEAR COLLEGES, prepared by CUPM. Presents listings of books and journals that should be in every college library. 106 pp. Paperbound.
List: $9.00. MAA Member: $6.50

THE CHAUVENET PAPERS. *A Collection of Prize Winning Expository Papers in Mathematics,* edited by James C. Abbott. Two-volumes of the collected prize winning Chauvenet Papers. Vol. 1—312 pp. Hardbound. Vol. 2—282 pp. Hardbound.
List: $21.00 each. MAA Member $16.00 each.
[Two volume sets
List: $36.00. MAA Member: $27.00.]

CRITICAL VARIABLES IN MATHEMATICS EDUCATION: *Findings from a Survey of the Empirical Literature,* by E. G. Begle. A joint publication of the MAA and the National Council of Teachers of Mathematics.
List: $8.00. MAA Member: $6.40.

A COMPENDIUM OF CUPM RECOMMENDATIONS. Volumes I and II. A collection of the major recommendations of the Committee on the Undergraduate Program in Mathematics. Two volumes. 756 pp. Hardbound.
List: $16.50. MAA Member: $12.00.

THE WILLIAM LOWELL, PUTNAM MATHEMATICAL COMPETITION: PROBLEMS AND SOLUTIONS—1938-1964. Compiled the R. E. Greenwood, A. M. Gleason, and L. M. Kelly. Contains problems and solutions to the first 25 Putnam Exams. 652 pp. Hardbound.
List: $35.00. MAA Member: $26.00.

THE MATHEMATICAL ASSOCIATION OF AMERICA: *Its First Fifty Years.* An historical perspective of the Association. 170 pp. Hardbound.
List: $10.00. MAA Member: $5.00.

FIFTY YEAR INDEX OF THE MATHEMATICS MAGAZINE, edited by Lynn A. Steen, and J. Arthur Seebach. Cumulative index of volumes 1-50. 163 pp. paperbound.
List: $10.00. MAA Member: $6.50.

INDEX OF THE AMERICAN MATHEMATICAL MONTHLY. Contains the tables of contents for each issue of volumes 1-80 as well as subject and author indices. 269 pp. Hardbound.
List: $19.00. MAA Member: $13.00.

PRIME-80. *Proceedings of a Conference on Prospects in Mathematics Education in the 1980's.* Included are the background of the conference and the recommendations that resulted.
84 pp. Paperbound. $3.50.

PROFESSIONAL OPPORTUNITIES IN THE MATHEMATICAL SCIENCES, Eleventh Edition. 1983. Designed for the student interested in a career in mathematics. 35 pp. Paperbound. $1.50 each. 95¢ for orders of five or more.

RECOMMENDATIONS ON A GENERAL MATHEMATICAL SCIENCES PROGRAM. Prepared by the Committee on the Undergraduate Program in Mathematics. (CUPM) 102 pp. Paperbound. $3.50 each

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order From:

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 90, Number 7 August-September 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

ITERATIVE PROCESSES IN ELEMENTARY GEOMETRY

GENG-ZHE CHANG* AND PHILIP J. DAVIS
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

1. Introduction. Certain constructions of elementary geometry, when iterated, give rise to figures which are often both of great visual appeal and whose mathematical analysis has attractive and difficult features. An example of such a figure is the "nested polygon." A planar polygon is given. Each side is divided into a fixed ratio and the points of division become the vertices of a second polygon. This construction is then iterated (Fig. 1).

An analysis of this problem and further references may be found in Berlekamp, Gilbert and Sinden [**2**], Davis [**4**]. An early solution of the problem presented by I. J. Schoenberg [**6**] pointed out the intimate relation to finite Fourier analysis, while Davis [**4**] placed the problem within the context of circulant matrices.

In this paper we take up three iterative constructions of a different sort, and present a mathematical analysis of them. The constructions undoubtedly suggest generalizations and we have found numerous difficulties in the way of complete solutions.

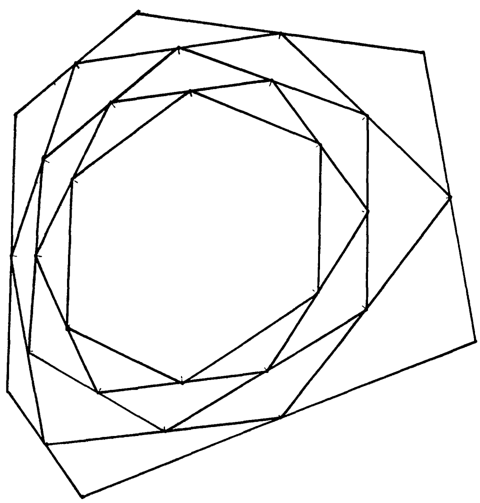


FIG. 1

2. Projected Simplexes. Let P be an arbitrary point lying in the triangle T_1. Determine second, third, ... triangles $T_2, T_3, \ldots$ as in Fig. 2. We wish to discuss the convergence properties of the figure. (Davis [**4**], page 3.) We present the solution for the n-simplex. Let $T_0, T_1, \ldots, T_n$ be the $n+1$ vertiçes of an n-simplex T: $(= T^{(0)})$. A point P inside the simplex can be expressed by

$$P = \lambda_0 T_0 + \lambda_1 T_1 + \cdots + \lambda_n T_n \tag{2.1}$$

Geng-zhe Chang is Associate Professor of Mathematics at the University of Science and Technology of China, Hefei, Anhui. Professor Chang's books include *Mathematical Foundations for Computing Aircraft Shapes* (1977), *Geometrical Problems by Complex Numbers* (1980).

Philip J. Davis is Professor of Applied Mathematics at Brown University. His extensive work in numerical analysis and applied mathematics includes the books *Interpolation and Approximation* (1963), *Numerical Integration* (with P. Rabinowitz, 1967), *The Schwarz Function and its Applications* (1974), *Circulant Matrices* (1979). Another book, *The Mathematical Experience*, coauthored with Professor Reuben Hersh, appeared recently.

*On leave from the Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, China.

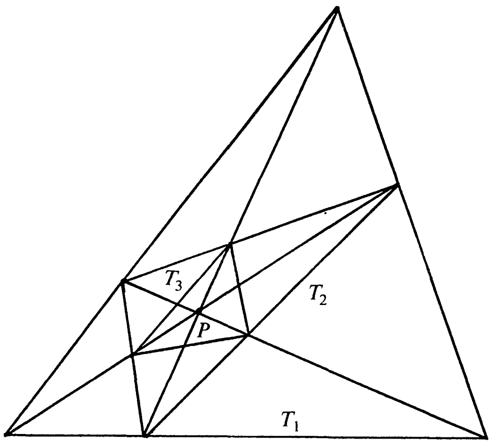


FIG. 2

where $0 < \lambda_i < 1$ and $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$. $(\lambda_0, \lambda_1, \ldots, \lambda_n)$ are called the barycentric coordinates of P with respect to the simplex.

Joining T_0 and P by a line segment, the intersection of this segment and the face opposite to T_0 is denoted by $T_0^{(1)}$ (Fig. 3). Since $T_0 = (1,0,0,\ldots,0)$, a typical point on the straight line determined by T_0 and P is

$$(1-t)T_0 + tP = \left[(1-t)1 + t\lambda_0, \ldots, \right].$$

Making the first barycentric coordinate zero, we see that t must be

$$t = \frac{1}{1-\lambda_0}.$$

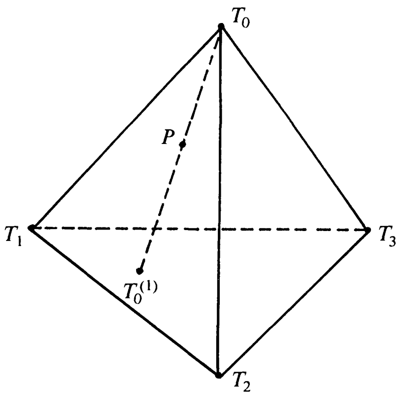


FIG. 3

Thus we get

$$T_0^{(1)} = \frac{-\lambda_0}{1-\lambda_0} T_0 + \frac{1}{1-\lambda_0} P = [0, \lambda_1, \lambda_2, \ldots, \lambda_n]/(1-\lambda_0).$$

$T_1^{(1)}, T_2^{(1)}, \ldots, T_n^{(1)}$ are similarly determined, and we have

$$T_1^{(1)} = [\lambda_0, 0, \lambda_2, \ldots, \lambda_n]/(1-\lambda_1),$$

$$\ldots$$

$$T_n^{(1)} = [\lambda_0, \lambda_1, \ldots, \lambda_{n-1}, 0]/(1-\lambda_n).$$

The $n+1$ equalities above can be rewritten in matrix notation

(2.2)

$$T^{(1)} \equiv \begin{bmatrix} T_0^{(1)} \\ T_1^{(1)} \\ \vdots \\ T_n^{(1)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{1-\lambda_0} & & & \\ & \frac{1}{1-\lambda_1} & & \\ & & \ddots & \\ & & & \frac{1}{1-\lambda_n} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ & & \cdots & & \\ 1 & 1 & 1 & \cdots & 1 \; 0 \end{bmatrix} \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} T_0 \\ T_1 \\ \vdots \\ T_n \end{bmatrix}.$$

The $T_i^{(1)}$ are the $(n+1)$ vertices of a simplex $T^{(1)}$.

Write $J =$ the $(n+1) \times (n+1)$ matrix, all of whose entries are 1, $D = \operatorname{diag}(\lambda_0, \lambda_1, \ldots, \lambda_n)$. Then $I - D = \operatorname{diag}(1-\lambda_0, 1-\lambda_1, \ldots, 1-\lambda_n)$ and $(I-D)^{-1} = \operatorname{diag}((1-\lambda_0)^{-1}, \ldots, (1-\lambda_n)^{-1})$. Thus (2.2) can be written in the form

(2.3) $$T^{(1)} = (I-D)^{-1}(J-I)DT.$$

Since $J^2 = (n+1)J$, it is easily verified that

(2.4) $$(J-I)^{-1} = \frac{1}{n}J - I.$$

Hence (2.3) may be inverted to yield

(2.5) $$T = D^{-1}\left(\frac{1}{n}J - I\right)(I-D)T^{(1)}.$$

Now

$$\begin{aligned} P &= (\lambda_0, \lambda_1, \ldots, \lambda_n)T \\ &= (\lambda_0, \lambda_1, \ldots, \lambda_n)D^{-1}\left(\frac{1}{n}J - I\right)(I-D)T^{(1)} \\ &= \frac{1}{n}(1-\lambda_0, 1-\lambda_1, \ldots, 1-\lambda_n)T^{(1)}. \end{aligned}$$

It follows that the barycentric coordinates of P with respect to $T^{(1)}$ are $\frac{1}{n}(1-\lambda_0, 1-\lambda_1, \ldots, 1-\lambda_n)$. Call them $\lambda_0^{(1)}, \ldots, \lambda_n^{(1)}$ (and those of $T = T^{(0)}$, $\lambda_i^{(0)}$) so that

(2.6) $$\lambda_i^{(1)} = \frac{1}{n}\left(1 - \lambda_i^{(0)}\right).$$

For the general iterative step, we have

(2.7) $$T^{(p+1)} = M_p T^{(p)},$$

where

(2.8) $$M_p = \left(I - D_p\right)^{-1}(J-I)D_p$$

with

$$D_p = \operatorname{diag}\left(\lambda_0^{(p)}, \ldots, \lambda_n^{(p)}\right). \tag{2.9}$$

Also, similarly to (2.6) we have $\lambda_i^{(p+1)} = \frac{1}{n}(1 - \lambda_i^{(p)})$ which can be written in matrix form as

$$D_{p+1} = \frac{1}{n}(I - D_p). \tag{2.10}$$

From (2.10), $(I - D_p)^{-1} = \frac{1}{n} D_{p+1}^{-1}$, so that

$$M_p = \frac{1}{n} D_{p+1}^{-1}(J - I) D_p. \tag{2.11}$$

From (2.7) it follows that

$$T^{(p)} = M_{p-1} M_{p-2} \cdots M_1 M_0 T. \tag{2.12}$$

Define $C = J - I$. Then from (2.12) and (2.11) we obtain

$$T^{(p)} = \frac{1}{n^p} D_p^{-1} C^p D_0 T. \tag{2.13}$$

By induction we obtain from (2.10)

$$D_p = \alpha_p I + \beta_p D_0, \quad p = 0, 1, \ldots, \tag{2.14}$$

where

$$\alpha_{p+1} = \frac{1}{n}(1 - \alpha_p), \quad \alpha_0 = 0, \quad \beta_{p+1} = -\frac{1}{n}\beta_p, \quad \beta_0 = 1. \tag{2.15}$$

The solutions of the difference equations (2.15) are

$$\alpha_p = \frac{1}{n+1} + \left(-\frac{1}{n+1}\right)\left(-\frac{1}{n}\right)^p, \qquad \beta_p = \left(-\frac{1}{n}\right)^p, \tag{2.16}$$

and

$$(n+1)\alpha_p + \beta_p = 1. \tag{2.16'}$$

Since the eigenvalues of J are $n+1, 0, 0, \ldots, 0$, those of $C = J - I$ are $n, -1, -1, \ldots, -1$. Thus, we have the diagonalization (see Davis [4], p. 72)

$$C = J - I = F^* \operatorname{diag}(n, -1, -1, \ldots, -1) F, \tag{2.17}$$

where F is the Fourier matrix of order $n+1$ and * designates the conjugate transpose. (The j, k element of F^* is $(n+1)^{-\frac{1}{2}} w^{(j-1)(k-1)}$, $w = \exp(2\pi i/(n+1))$.) From (2.17) and (2.13), we have

$$T^{(p)} = D_p^{-1} F^* \operatorname{diag}\left(1, \left(-\frac{1}{n}\right)^p, \ldots, \left(-\frac{1}{n}\right)^p\right) F D_0 T. \tag{2.18}$$

We may now allow $p \to \infty$ in (2.18). From (2.16), for $n \geqslant 2$,

$$\lim_{p \to \infty} \alpha_p = \frac{1}{n+1}, \quad \lim_{p \to \infty} \beta_p = 0,$$

so that from (2.14),

$$\lim_{p \to \infty} D_p = \frac{1}{n+1} I \text{ and } \lim_{p \to \infty} D_p^{-1} = (n+1) I.$$

Thus, from (2.18),

$$T^{(\infty)} = (n+1) F^* \operatorname{diag}(1, 0, 0, \ldots, 0) F D_0. \tag{2.19}$$

Since $F^*\text{diag}(1,0,\ldots,0)F = \frac{1}{n+1}J$,

$$T^{(\infty)} = JD_0\begin{pmatrix} T_0 \\ \vdots \\ T_n \end{pmatrix} = J\begin{pmatrix} \lambda_0 T_0 \\ \vdots \\ \lambda_n T_n \end{pmatrix} = \begin{pmatrix} P \\ \vdots \\ P \end{pmatrix} \equiv \mathbb{P}$$

and this verifies what is visually obvious, namely, the convergence of the vertices of the simplex to the point P.

It is of interest to discuss the volumes of the simplexes. Let $\Delta^{(p)}$ and Δ designate the volumes of the simplexes $T^{(p)}$ and T respectively. From (2.18) it follows that

$$\begin{aligned}\frac{\Delta^{(p)}}{\Delta} &= \left|\det\left(D_p^{-1}F^*\text{diag}\left(1,\left(-\frac{1}{n}\right)^p,\ldots,\left(-\frac{1}{n}\right)^p\right)FD_0\right|\\ &= \left(\frac{1}{n}\right)^{pn}\lambda_0\lambda_1\cdots\lambda_n/\det D_p.\end{aligned}$$

(See, e.g., Benson [1], p. 188.)

Allowing $p \to \infty$, we obtain

$$\lim_{p\to\infty}(n^n)^p\frac{\Delta^{(p)}}{\Delta} = (n+1)^{n+1}\lambda_0\lambda_1\cdots\lambda_n. \tag{2.20}$$

To discuss the rate of convergence, we may proceed as follows:

$$\text{diag}(1,\beta_p,\ldots,\beta_p) = \text{diag}(\beta_p,\ldots,\beta_p) + \text{diag}(1-\beta_p,0,0,\ldots,0).$$

Hence,

$$F^*\text{diag}(1,\beta_p,\ldots,\beta_p)F = \beta_p I + \frac{1-\beta_p}{n+1}J.$$

Thus, from (2.18),

$$\begin{aligned}T^{(p)} &= D_p^{-1}\left(\beta_p I + \frac{1-\beta_p}{n+1}J\right)D_0 T\\ &= D_p^{-1}\left(\beta_p D_0 T + \frac{1-\beta_p}{n+1}\mathbb{P}\right).\end{aligned} \tag{2.21}$$

In view of (2.14), (2.16′),

$$D_p = \frac{(1-\beta_p)}{n+1}I + \beta_p D_0. \tag{2.22}$$

Combining this with (2.21) yields

$$\frac{1}{\beta_p}(T^{(p)} - \mathbb{P}) = D_p^{-1}D_0(T - \mathbb{P});$$

hence

THEOREM 2.1:

$$\lim_{p\to\infty}\frac{1}{\beta_p}(T^{(p)} - \mathbb{P}) = (n+1)D_0(T-\mathbb{P}). \tag{2.23}$$

Thus, the vertices of the simplex $T^{(p)}$ converge to P with geometric rapidity $(-1/n)$, and the right-hand side of (2.23) identifies the leading coefficient in the asymptotic expansion.

3. Another Generalization of the Plane Case. In this section we consider a certain iterative

process which starts from a given triangle and which generalizes the plane case in a different way. It is particularly convenient in this case to place the triangle in the complex plane. We denote a point in the plane and its complex coordinates both by the same letter. For example, the triangle abc represents a triangle with the vertices whose complex coordinates are a, b and c.

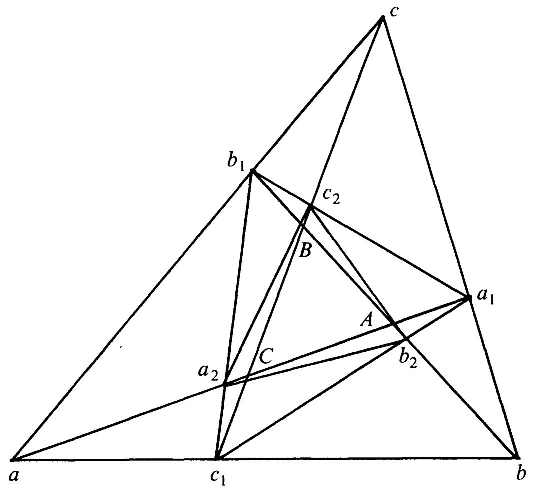


FIG. 4

Assume that a triangle abc is arbitrarily given and that a_1, b_1, c_1 are points on its sides bc, ca, ab respectively (Fig. 4). The intersection of the line segments aa_1 and b_1c_1 is denoted by a_2. Symmetrically we shall use b_2, c_2 for the other two intersections. Do the same thing to the triangle $a_1b_1c_1$ and three points a_2, b_2, c_2 on each side, a triangle $a_3b_3c_3$ will be obtained. Generally, starting from the triangle $a_{n-1}b_{n-1}c_{n-1}$ and three points a_n, b_n, c_n on its sides, a triangle $a_{n+1}b_{n+1}c_{n+1}$ will be constructed by the same method. The following questions arise:

(1) To what points will a_n, b_n, c_n converge as $n \to \infty$?

(2) How fast is the convergence?

Assume that

$$\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 & 1-\alpha & \alpha \\ \beta & 0 & 1-\beta \\ 1-\gamma & \gamma & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{3.1}$$

where $0 \leqslant \alpha, \beta, \gamma \leqslant 1$. The numbers α, β, γ are completely determined once a_1, b_1, c_1 are given. To solve these two problems, we must express a_n, b_n, c_n in terms of a, b, c and α, β, γ. We begin with the case $n = 1$. Since a_2 is on both line segments b_1c_1 and aa_1, we can write a_2 as

$$(1-\alpha_1)b_1 + \alpha_1 c_1 = (1-\mu)a + \mu a_1 \tag{3.2}$$

where $0 \leqslant \alpha_1, \mu \leqslant 1$. α_1 can be determined as follows. Inserting (3.1) into both sides of (3.2) and equating the coefficients of b, c, respectively, we obtain

$$\begin{gathered} \alpha_1\gamma = \mu(1-\alpha), \\ (1-\alpha_1)(1-\beta) = \alpha\mu. \end{gathered} \tag{3.3}$$

Solving α_1 from (3.3), we get

$$\alpha_1 = \frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)+\alpha\gamma}.$$

After similar manipulations have been applied to b_2 and c_2, we get

$$\begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 & 1-\alpha_1 & \alpha_1 \\ \beta_1 & 0 & 1-\beta_1 \\ 1-\gamma_1 & \gamma_1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}, \tag{3.4}$$

where

$$\begin{aligned} \alpha_1 &= \frac{(1-\alpha)(1-\beta)}{(1-\alpha)(1-\beta)+\alpha\gamma}, \\ \beta_1 &= \frac{(1-\beta)(1-\gamma)}{(1-\beta)(1-\gamma)+\beta\alpha}, \\ \gamma_1 &= \frac{(1-\gamma)(1-\alpha)}{(1-\gamma)(1-\alpha)+\gamma\beta}. \end{aligned} \tag{3.5}$$

Incidentally, in the above we have employed a standard method for finding the intersection of two line segments whose end-points are known in complex coordinates.

When we proceed from the triangle $a_nb_nc_n$ to the triangle $a_{n+1}b_{n+1}c_{n+1}$, no additional effort need be made except to raise each subscript by $n-1$ in (3.4) and (3.5):

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1-\alpha_n & \alpha_n \\ \beta_n & 0 & 1-\beta_n \\ 1-\gamma_n & \gamma_n & 0 \end{bmatrix} \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix}, \tag{3.6}$$

where

$$\begin{aligned} \alpha_n &= \frac{(1-\alpha_{n-1})(1-\beta_{n-1})}{(1-\alpha_{n-1})(1-\beta_{n-1})+\alpha_{n-1}\gamma_{n-1}}, \\ \beta_n &= \frac{(1-\beta_{n-1})(1-\gamma_{n-1})}{(1-\beta_{n-1})(1-\gamma_{n-1})+\beta_{n-1}\alpha_{n-1}}, \\ \gamma_n &= \frac{(1-\gamma_{n-1})(1-\alpha_{n-1})}{(1-\gamma_{n-1})(1-\alpha_{n-1})+\gamma_{n-1}\beta_{n-1}}, \end{aligned} \tag{3.7}$$

$n = 1,2,3,\ldots$, with $\alpha_0 = \alpha$, $\beta_0 = \beta$, $\gamma_0 = \gamma$.

Let A, B, C be the intersections two by two cyclically of the segments aa_1, bb_1, cc_1. By the general method shown above we have

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\gamma+\beta\gamma} & & \\ & \frac{1}{1-\alpha+\gamma\alpha} & \\ & & \frac{1}{1-\beta+\alpha\beta} \end{bmatrix} \times \begin{bmatrix} \beta(1-\gamma) & \beta\gamma & (1-\beta)(1-\gamma) \\ (1-\gamma)(1-\alpha) & \gamma(1-\alpha) & \gamma\alpha \\ \alpha\beta & (1-\alpha)(1-\beta) & \alpha(1-\beta) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \tag{3.8}$$

Geometric constructions suggest to us intuitively that the sequence of triangles $a_n b_n c_n$ will converge to the triangle ABC in a certain manner.

Denote the areas of the triangle ABC and the triangle abc by $\Delta(ABC)$ and $\Delta(abc)$ respectively. From (3.8) it follows that

$$\Delta(ABC) = \frac{\Delta(abc)}{(1-\gamma+\beta\gamma)(1-\alpha+\gamma\alpha)(1-\beta+\alpha\beta)} \times \begin{vmatrix} \beta(1-\gamma) & \beta\gamma & (1-\beta)(1-\gamma) \\ (1-\gamma)(1-\alpha) & \gamma(1-\alpha) & \gamma\alpha \\ \alpha\beta & (1-\alpha)(1-\beta) & \alpha(1-\beta) \end{vmatrix},$$

i.e.,

$$\Delta(ABC) = \frac{[\alpha\beta\gamma - (1-\alpha)(1-\beta)(1-\gamma)]^2}{(1-\gamma+\beta\gamma)(1-\alpha+\gamma\alpha)(1-\beta+\alpha\beta)} \Delta(abc). \tag{3.9}$$

This formula was discovered by E. J. Routh in 1896 (see *Introduction to Geometry* by Coxeter, p. 211). Since $A = B = C$ iff $\Delta(ABC) = 0$, hence from (3.9), $A = B = C$ if and only if

$$\alpha\beta\gamma = (1-\alpha)(1-\beta)(1-\gamma). \tag{3.10}$$

This is the famous theorem of Ceva and the lines aa_1, bb_1, cc_1 are often called *Cevians* of the triangle abc (see *Introduction to Geometry*, p. 220).

If (3.10) holds, then aa_1, bb_1, cc_1 meet at the same point and the problems are reduced to the planar case discussed carefully in the previous section. Hence we assume $\alpha\beta\gamma \neq (1-\alpha)(1-\beta)(1-\gamma)$. In this case we can invert (3.8) to get

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{\alpha\beta\gamma - (1-\alpha)(1-\beta)(1-\gamma)} \begin{bmatrix} 0 & -(1-\beta) & \gamma \\ \alpha & 0 & -(1-\gamma) \\ -(1-\alpha) & \beta & 0 \end{bmatrix} \times \begin{bmatrix} 1-\gamma+\beta\gamma & & \\ & 1-\alpha+\gamma\alpha & \\ & & 1-\beta+\alpha\beta \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix}. \tag{3.11}$$

It seems to us very complicated to deal with the general case, so we assume that $\alpha = \beta = \gamma \neq \frac{1}{2}$. In this case we have $\alpha_n = \beta_n = \gamma_n$ ($n = 0, 1, 2, \cdots$), and (3.6), (3.7), (3.11) are reduced respectively to the following simpler forms

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{bmatrix} = [(1-\alpha_n)\Pi + \alpha_n \Pi^2] \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} \tag{3.12}$$

where Π is the simple circulant matrix:

$$\Pi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix};$$

$$\alpha_n = \frac{(1-\alpha_{n-1})^2}{(1-\alpha_{n-1})^2 + \alpha_{n-1}^2} \tag{3.13}$$

for $n = 1, 2, 3, \cdots$ and

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{1-2\alpha} [(1-\alpha)\Pi - \alpha\Pi^2] \begin{bmatrix} A \\ B \\ C \end{bmatrix}. \tag{3.14}$$

We need the following lemmas:

LEMMA 1. *We have*

$$\alpha_n = \begin{cases} \dfrac{\alpha^{2^n}}{\alpha^{2^n} + (1-\alpha)^{2^n}}, & n \text{ even}, \\ \dfrac{(1-\alpha)^{2^n}}{\alpha^{2^n} + (1-\alpha)^{2^n}}, & n \text{ odd}. \end{cases} \tag{3.15}$$

The proof is omitted. It is a straightforward mathematical induction.

It will be more convenient to express a_n, b_n, c_n in terms of A, B, C. To make the expressions neater, we introduce the following 3×3 matrix

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Note that $Q^n = I$, the 3×3 identity matrix, when n is even and $Q^n = Q$ when n is odd, and that

$$Q\Pi Q = \Pi^2, \quad Q\Pi^2 Q = \Pi. \tag{3.16}$$

LEMMA 2. *Let*

$$\lambda_n = \left(\frac{\alpha}{1-\alpha}\right)^{2^n}, \tag{3.17}$$

for $n = 0, 1, 2, \dots$; *then we have*

$$\begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix} = \frac{1}{1-\lambda_n} Q^n(\Pi - \lambda_n \Pi^2) Q^n \begin{bmatrix} A \\ B \\ C \end{bmatrix}. \tag{3.18}$$

Proof. By induction.

For $n = 0$, (3.18) reduces to (3.14).

Assume that (3.18) is true for even integers n. Then by (3.12) and (3.18) we have

$$\begin{aligned} \begin{bmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{bmatrix} &= \left[(1-\alpha_n)\Pi + \alpha_n \Pi^2\right] \frac{1}{1-\lambda_n}(\Pi - \lambda_n \Pi^2) \begin{bmatrix} A \\ B \\ C \end{bmatrix} \\ &= \frac{1}{1-\lambda_n}\left[-\alpha_n \lambda_n \Pi + (1-\alpha_n)\Pi^2 + (\alpha_n - \lambda_n(1-\alpha_n))I\right] \begin{bmatrix} A \\ B \\ C \end{bmatrix}. \end{aligned}$$

Since for even n from (3.17) and (3.15) we see that

$$\lambda_n = \frac{\alpha_n}{1-\alpha_n},$$

thus

$$1 - \alpha_n = \frac{1}{1+\lambda_n}, \qquad \alpha_n \lambda_n = \frac{\lambda_n^2}{1+\lambda_n},$$

hence

$$\begin{bmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \end{bmatrix} = \frac{1}{(1-\lambda_n)(1+\lambda_n)}(-\lambda_n^2 \Pi + \Pi^2) \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

$$= \frac{1}{1-\lambda_n^2} Q(\Pi - \lambda_n^2 \Pi^2) Q \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

$$= \frac{1}{1-\lambda_{n+1}} Q^{n+1}(\Pi - \lambda_{n+1}\Pi^2) Q^{n+1} \begin{bmatrix} A \\ B \\ C \end{bmatrix}.$$

Similarly, if we assume that (3.18) is true for odd n, we can also prove (3.18) is true for $n + 1$. This completes the induction.

From (3.18) we have

$$a_n = \begin{cases} (B - \lambda_n C)/(1 - \lambda_n), & \text{for } n \text{ even} \\ (C - \lambda_n B)/(1 - \lambda_n), & \text{for } n \text{ odd.} \end{cases} \tag{3.19}$$

Without loss of generality we assume that $0 < \alpha < \frac{1}{2}$; in this case $0 < \frac{\alpha}{1-\alpha} < 1$. Hence from (3.19) it follows that

$$a_n = \begin{cases} B + \lambda_n(B - C) + O(\lambda_n^2), & n \text{ even} \\ C + \lambda_n(C - B) + O(\lambda_n^2), & n \text{ odd.} \end{cases} \tag{3.20}$$

Since $\lambda_n \to 0$ as $n \to \infty$, then by (3.20) we have the convergence of the sequences of the vertices, as shown in the following theorem.

THEOREM 3.1. *For* $0 < \alpha < \frac{1}{2}$, *we have*

$$\lim a_n = \begin{cases} B, & \text{even } n \to \infty \\ C, & \text{odd } n \to \infty \end{cases}$$

$$\lim b_n = \begin{cases} C, & \text{even } n \to \infty \\ A, & \text{odd } n \to \infty \end{cases}$$

$$\lim c_n = \begin{cases} A, & \text{even } n \to \infty \\ B, & \text{odd } n \to \infty. \end{cases}$$

From (3.18) we have

$$\Delta(a_n b_n c_n) = \frac{1 - \lambda_n^3}{(1 - \lambda_n)^3} \Delta(ABC) = \frac{1 + \lambda_n + \lambda_n^2}{(1 - \lambda_n)^2} \Delta(ABC)$$
$$= \Delta(ABC) + 3\lambda_n \Delta(ABC) + \text{smaller terms.}$$

Hence problem (2) is answered by

THEOREM 3.2. *Assume that* $0 < \alpha < \frac{1}{2}$, *then*

$$0 < \Delta(a_n b_n c_n) - \Delta(ABC) = O\left[\left(\frac{\alpha}{1-\alpha}\right)^{2^n}\right] \tag{3.21}$$

as $n \to \infty$.

The planar case of Section 2 and Section 3 differ in that in Section 2 the cevians of the starting triangle are assumed to be concurrent while in Section 3 they are not. The interesting and surprising thing about the asymptotics of convergence is that in the first case, the convergence is geometric, while in the second, it is supergeometric. (Numerical analysts would say that the convergence is linear in the first and quadratic in the second cases.)

This difference is reflected in the fact that if one draws the figures with a ruler and pencil, in the first case many iterations may be drawn, but in the second, only a few. One might say therefore that the difference in the rates of convergence can be intuited visually.

References

1. Russell V. Benson, Euclidean Geometry and Convexity, McGraw-Hill, New York,1966.
2. E. R. Berlekamp, E. N. Gilbert, and F. W. Sinden, A polygon problem, this MONTHLY, vol. 72 (1965) 233–241.
3. H. S. M. Coxeter, Introduction to Geometry, John Wiley, New York, 1961.
4. P. J. Davis, Circulant Matrices, John Wiley, New York, 1979.
5. E. J. Routh, A Treatise on Analytical Statics, with Numerous Examples, Cambridge University Press, 1896.
6. I. J. Schoenberg, The finite Fourier series and elementary geometry, this MONTHLY, vol. 57 (1950) 390–404.

A STATISTICAL ANALYSIS OF CASINO BLACKJACK

MARTIN H. MILLMAN
Department of Mathematics, William Paterson College, Wayne, NJ 07470

Introduction. A recent (3/82) ruling of the New Jersey Supreme Court, preventing legal gambling casinos from expelling "skilled" blackjack players, prompted the initial interest for undertaking this study. With a view toward determining what advantage a player (or group of players) using correct strategy might obtain over the "house" in casino blackjack, a Vax 11/780 computer was programmed to simulate the game. The model for this simulation was one "player" versus the "dealer," with the computer executing both roles. The program shuffled and dealt the cards, placed the "player's" bets, played the "player's" and "dealer's" hands, settled the bets, kept track of the "player's" capital and, in general, followed actual casino rules* and procedures in every essential detail. Separate experiments were performed for the Las Vegas and the Atlantic City versions of the game. In Las Vegas blackjack, four standard decks of cards are employed, three of which are dealt out before reshuffling. In Atlantic City six decks are employed, four of which are dealt out before reshuffling**. In both cases, the rules governing play of the hands are identical.

In order to investigate the expected win for a player using a correct strategy, a "card-counting" strategy, developed by Baldwin, et al. [**1**], Thorp [**2**], [**3**], Braun [**4**] and others, was incorporated in the program. The player's bets and playing decisions were then made by the computer in accordance with this strategy. The strategy, as will be seen presently, yields modest wins for the player under both versions of the rules.

The dealer is allowed no playing decisions in casino blackjack. The dealer's hands were executed according to the fixed rule: "hit" totals of 16 or less and "stand" on totals of 17 or more.

The Player Strategy. The program provides the "player" with a winning strategy that consists

---

Martin H. Millman received his B.S. at City College of New York (1962) and his Ph.D. at New York University (1968) under J. B. Keller. He has been a postdoctoral fellow at the Courant Institute of Mathematical Sciences, has taught at New York University, New Jersey Institute of Technology, and Fairleigh Dickinson University and is currently a member of the Mathematics Department at William Paterson College, Wayne, New Jersey. As a result of the research reported here, his interest in blackjack has been limited to an occasional game with his little daughter, Tasha. Other interests include paddleball, chinese checkers, and computer graphics.

*Complete rules of casino blackjack may be found in [**2**]. Our simulation allowed the player to double down on split pairs, but did not allow resplitting of split pairs. This is standard practice in Atlantic City casinos and common in Las Vegas.

**A crucial aspect of any simulation of this kind is the ability of the random number generator to supply "pseudorandom" sequences of integers. In this experiment such sequences were used to produce "shuffled" decks of cards. The generator employed was of the linear congruential type. It was χ^2-tested and found acceptable as a source of sample integers from a uniform distribution of integers in the interval [1, 312].

References

1. Russell V. Benson, Euclidean Geometry and Convexity, McGraw-Hill, New York, 1966.
2. E. R. Berlekamp, E. N. Gilbert, and F. W. Sinden, A polygon problem, this MONTHLY, vol. 72 (1965) 233–241.
3. H. S. M. Coxeter, Introduction to Geometry, John Wiley, New York, 1961.
4. P. J. Davis, Circulant Matrices, John Wiley, New York, 1979.
5. E. J. Routh, A Treatise on Analytical Statics, with Numerous Examples, Cambridge University Press, 1896.
6. I. J. Schoenberg, The finite Fourier series and elementary geometry, this MONTHLY, vol. 57 (1950) 390–404.

A STATISTICAL ANALYSIS OF CASINO BLACKJACK

MARTIN H. MILLMAN
Department of Mathematics, William Paterson College, Wayne, NJ 07470

Introduction. A recent (3/82) ruling of the New Jersey Supreme Court, preventing legal gambling casinos from expelling "skilled" blackjack players, prompted the initial interest for undertaking this study. With a view toward determining what advantage a player (or group of players) using correct strategy might obtain over the "house" in casino blackjack, a Vax 11/780 computer was programmed to simulate the game. The model for this simulation was one "player" versus the "dealer," with the computer executing both roles. The program shuffled and dealt the cards, placed the "player's" bets, played the "player's" and "dealer's" hands, settled the bets, kept track of the "player's" capital and, in general, followed actual casino rules* and procedures in every essential detail. Separate experiments were performed for the Las Vegas and the Atlantic City versions of the game. In Las Vegas blackjack, four standard decks of cards are employed, three of which are dealt out before reshuffling. In Atlantic City six decks are employed, four of which are dealt out before reshuffling**. In both cases, the rules governing play of the hands are identical.

In order to investigate the expected win for a player using a correct strategy, a "card-counting" strategy, developed by Baldwin, et al. [**1**], Thorp [**2**], [**3**], Braun [**4**] and others, was incorporated in the program. The player's bets and playing decisions were then made by the computer in accordance with this strategy. The strategy, as will be seen presently, yields modest wins for the player under both versions of the rules.

The dealer is allowed no playing decisions in casino blackjack. The dealer's hands were executed according to the fixed rule: "hit" totals of 16 or less and "stand" on totals of 17 or more.

The Player Strategy. The program provides the "player" with a winning strategy that consists

Martin H. Millman received his B.S. at City College of New York (1962) and his Ph.D. at New York University (1968) under J. B. Keller. He has been a postdoctoral fellow at the Courant Institute of Mathematical Sciences, has taught at New York University, New Jersey Institute of Technology, and Fairleigh Dickinson University and is currently a member of the Mathematics Department at William Paterson College, Wayne, New Jersey. As a result of the research reported here, his interest in blackjack has been limited to an occasional game with his little daughter, Tasha. Other interests include paddleball, chinese checkers, and computer graphics.

*Complete rules of casino blackjack may be found in [**2**]. Our simulation allowed the player to double down on split pairs, but did not allow resplitting of split pairs. This is standard practice in Atlantic City casinos and common in Las Vegas.

**A crucial aspect of any simulation of this kind is the ability of the random number generator to supply "pseudorandom" sequences of integers. In this experiment such sequences were used to produce "shuffled" decks of cards. The generator employed was of the linear congruential type. It was χ^2-tested and found acceptable as a source of sample integers from a uniform distribution of integers in the interval [1, 312].

of two components: the "truecount," which determines the player's bets from hand to hand, and the method for playing the hands. Briefly, the player assigns the "count-value" $+1$ to each 2, 3, 4, 5 and 6-point card, the count-value 0 to each 7, 8 and 9-point card, and the count-value -1 to each 10-point card and each ace. The player then keeps a running total of these values as cards appear during the game. Whenever shuffling occurs, this running total is reset to zero. Immediately before each betting opportunity, the player computes the "truecount" by dividing the momentary value of the running total by the number of thus far unused decks of cards. The truecount measures the "concentration" of high cards (tens and aces) in the cards remaining to be played before reshuffle. It is therefore proportional to the probability that the next card dealt will be a high one. Braun has shown that the player's expectation on the next hand is approximately proportional to (truecount-1), if the hand is played correctly. The strategy exploits this fact by prescribing a minimum bet on the next hand when the truecount is less than $+2$, and a maximum bet when the truecount equals or exceeds $+2$. The minimum bet in our experiment was taken to be one "chip" while the maximum bet was fixed at 20 "chips." (The monetary value of one chip is immaterial in our statistical analysis.) It is of interest to note that variations of bet size are essential for the success of the strategy. If, for example, the same bet is made on every hand, then the player's rate of profit becomes so small compared to his risk, that the game is of no interest. By contrast, our 1-20 bet variation gives the player a rate of profit that approaches the maximum rate attainable by the strategy.

The method of playing the hand employed by our hypothetical player is summarized in matrix form in Table I. The row labels running down the left side of the table represent all possible player hands. The third group of hands, A2 through A10, represents the so-called "soft" hands. (Here, as elsewhere in the table, A stands for an ace.) These hands contain at least one ace and total eleven or less when the aces are counted as ones. They cannot bust, even if "hit" with a 10-point card. The number of cards in a hand is immaterial. For example, A8 represents hands such as A-8, A-5-3, or even A-2-3-A-2.

The second group of hands in Table I, 2-2 through A-A, represents two-card hands that are pairs. Finally, the first group, 5 through 21, represents hands, having the given totals, which are neither "soft" nor pairs. For example, 15 represents hands like 10-5, 7-6-2 or 9-5-A.

The column labels running across the top of Table I represent all possible dealer "up (exposed)-cards." The matrix entries determine the player's action for each possible combination of player hand and exposed dealer card. Possible player actions are abbreviated as follows: H = "hit," S = "stand," D = "double," P = "split a pair." Thus, for example, the player will stand with 5-3-4 if the dealer shows a 6 and split 7-7 if the dealer shows a 2.

In addition to the strategy already outlined, our player also made the optional "insurance" bet (see [**2**]) when the dealer's up-card was an ace if the truecount exceeded $+3$. Finally, if the truecount at the conclusion of a hand was negative, our player "left the table" to resume playing with freshly-shuffled cards elsewhere in the hypothetical casino. This tactic greatly reduced the proportion of hands played at negative expectation and thus increased the player's net profit significantly. It was simulated in the program by having the computer shuffle and begin dealing anew whenever the truecount at the conclusion of a hand was negative.

More sophisticated playing strategies exist than the one employed here. The principal improvement that can be introduced is to vary the playing decisions, as well as the bet size, as the truecount fluctuates. Braun [**4**] has in fact computed the optimal playing strategy for all practical values of truecount by simulation methods. While such improved strategies do result in slightly larger profits, the great burden they place on a human player's memory, concentration and accuracy probably renders them uneconomical. We have therefore utilized the simpler strategy discussed above.

Data and Results. The statistical parameters of a "game" consisting of twenty-thousand consecutive hands of casino blackjack, in which the player employs our strategy, were investi-

TABLE I. Playing strategy.

	2	3	4	5	6	7	8	9	10	A
5	H	H	H	H	H	H	H	H	H	H
6	H	H	H	H	H	H	H	H	H	H
7	H	H	H	H	H	H	H	H	H	H
8	H	H	H	H	H	H	H	H	H	H
9	H	D	D	D	D	H	H	H	H	H
10	D	D	D	D	D	D	D	D	H	H
11	D	D	D	D	D	D	D	D	D	H
12	H	H	S	S	S	H	H	H	H	H
13	S	S	S	S	S	H	H	H	H	H
14	S	S	S	S	S	H	H	H	H	H
15	S	S	S	S	S	H	H	H	H	H
16	S	S	S	S	S	H	H	H	S	H
17	S	S	S	S	S	S	S	S	S	S
18	S	S	S	S	S	S	S	S	S	S
19	S	S	S	S	S	S	S	S	S	S
20	S	S	S	S	S	S	S	S	S	S
21	S	S	S	S	S	S	S	S	S	S
2-2	P	P	P	P	P	P	H	H	H	H
3-3	P	P	P	P	P	P	H	H	H	H
4-4	H	H	H	H	H	H	H	H	H	H
5-5	D	D	D	D	D	D	D	D	H	H
6-6	H	P	P	P	P	H	H	H	H	H
7-7	P	P	P	P	P	P	H	H	H	H
8-8	P	P	P	P	P	P	P	P	P	P
9-9	P	P	P	P	P	S	P	P	S	S
10-10	S	S	S	S	S	S	S	S	S	S
A-A	P	P	P	P	P	P	P	P	P	P
A2	H	H	H	D	D	H	H	H	H	H
A3	H	H	H	D	D	H	H	H	H	H
A4	H	H	D	D	D	H	H	H	H	H
A5	H	H	D	D	D	H	H	H	H	H
A6	H	D	D	D	D	H	H	H	H	H
A7	S	D	D	D	D	S	S	H	H	S
A8	S	S	S	S	S	S	S	S	S	S
A9	S	S	S	S	S	S	S	S	S	S
A10	S	S	S	S	S	S	S	S	S	S

gated. Data were collected for samples consisting of 600 trials of twenty thousand hands each for the Las Vegas case and 400 trials of twenty thousand hands each for the Atlantic City case. In total, a random sample of twenty million hands was executed during nearly 21 (no pun) hours of computing time.

The probability that the player's final capital exceeds his initial capital after a game consisting of N hands is an increasing function of N. Hereafter we refer to this probability simply as the "win probability," denoted by P_w. The rather large value $N = 20{,}000$ was selected for this experiment in order to insure interestingly large values of P_w under both versions of the rules. Practically speaking, the actual playing of twenty thousand hands in a casino would consume some 300 to 400 human hours.

Table II gives a summary of the sample data. The first column reports the proportion of trials in the sample that resulted in a net win (of any amount) for the player. The second column reports the mean number of chips won, averaged over all the trials in the sample and rounded to the nearest whole chip. The third column then gives the sample standard deviation of these wins, again rounded to an integer.

We define the player's "edge" in a single twenty-thousand-hand trial to be the net number of

chips won divided by the total number of chips bet. The "edge" is therefore the player's rate of profit on his investment. The fourth column of the table reports the mean edge, averaged over all the trials in the sample.

Finally, the player was allowed unlimited credit in each independent trial and the worst state of his funds during each trial was tracked by the computer. At the conclusion of each trial the computer updated the largest single-trial deficit yet encountered. The largest single-trial deficit in the entire sample is reported in the last column. Thus, for example, our Las Vegas player had accumulated a net loss of 4961 chips at some point during his most distressing trial.

TABLE II. Sample data.

	Proportion of Wins	Mean Win (chips)	Standard Deviation (chips)	Mean Player Edge (%)	Largest Player Deficit (chips)
Las Vegas	.87	2084	1786	1.35	4961
Atlantic City	.79	1108	1471	0.91	3621

The quantities in each column of Table II serve as estimators for the quantities in the corresponding columns of Table III. The first three columns of Table III give 95% confidence intervals for the win probability P_w, the expected win μ, and the standard deviation σ respectively for our twenty-thousand-hand game. All chip-values are again rounded to the nearest integer.

The fourth column of Table III gives a very close estimate of the player's expected edge. The figures in the last column of the table are to be interpreted as rough statistical solutions of one aspect of the classical "gambler's ruin" problem. This problem recognizes the fact that although the player may be capable of winning in the long run, he may nevertheless be ruined (lose all his capital) in one of the large negative fluctuations that inevitably occur. In order to avoid this unhappy possibility, the player needs to know how much initial capital he requires to survive till the end of the game. Thus, for example, 4962 chips is the "required initial capital" for the Las Vegas game in the sense that a player possessing such capital (and employing our strategy) has zero experimental probability of going broke*.

Table III. Parameters for the twenty-thousand hand game.

	Win Probability	Expected Win (chips)	Standard deviation (chips)	Expected Player Edge (%)	Required Initial Capital (chips)
Las Vegas	$.84 \leqslant P_w \leqslant .90$	$1944 \leqslant \mu \leqslant 2227$	$1690 \leqslant \sigma \leqslant 1893$	1.35	4962
Atlantic City	$.75 \leqslant P_w \leqslant .83$	$964 \leqslant \mu \leqslant 1252$	$1375 \leqslant \sigma \leqslant 1580$	0.91	3622

Fig. 1 is a histogram representation of the 600-trial Las Vegas experiment. It shows the number of trials that fell within each profit interval of width 200, from -5000 to $+8000$ chips. For

*These "required capital" figures are merely rough estimates of the initial capital required to participate in the game with zero theoretical probability of going broke. There is over a 98% chance, though, that each exceeds the top 1% percentile of the underlying theoretical distribution of maximum deficit per trial. Of more practical value would have been the experimental distribution of maximum deficit. The top 5% percentile of this distribution, for example, would have been a rather accurate estimate of the capital necessary to play with a 5% probability of ruin. Regrettably, these more accurate percentile figures could not be obtained due to severe limitations on the amount of computer memory ("disk quota") that was available to the author. Thus, only the inaccurate 0% percentile figures quoted in Table III could be printed out. This omission might be remedied in a future rerun of the experiment.

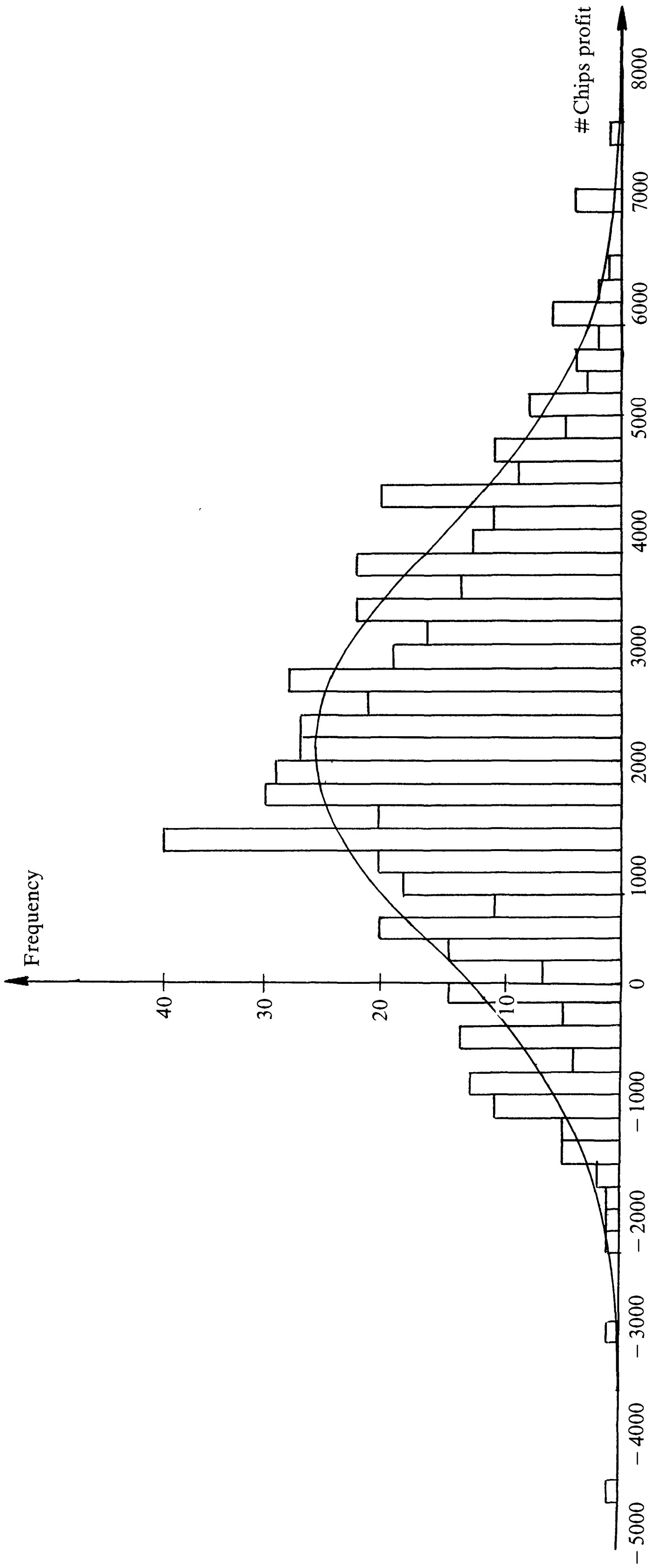

Frequency
40
30
20
10
0
Chips profit
−5000
−4000
−3000
−2000
−1000
0
1000
2000
3000
4000
5000
6000
7000
8000

FIG. 1.

example, the histogram shows that 40 of the 600 trials resulted in wins of between 1200 and 1400 chips, while only 5 trials resulted in losses of between 1200 and 1400 chips.

The total shaded area is equal to 120,000 chip-occurrences and 87% of this area lies to the right of the origin. A normal curve, normalized to area 120,000, having the same mean and standard deviation as the histogram ($\mu = 2084$, $\sigma = 1786$), is also shown in Fig. 1. The resemblance between histogram and curve supports the hypothesis that the theoretical distribution of player profits is approximately normal. That is, a χ^2 goodness-of-fit test accepts our sample as a random one from the normal distribution shown. Approximate normality of the theoretical distribution is at first somewhat surprising in view of the player's varying bet size and the fact that consecutive hands of blackjack are not strictly independent events. However, a more detailed analysis does lead one to expect approximate normality.

Conclusions. Table III shows that when the same strategy is employed by the player in both the 4-deck and 6-deck games, the expected win is about twice as great in the 4-deck case, while the initial capital required is, roughly, only about one and a half times as great. Furthermore, the relative fluctuation of the player's funds is less in the 4-deck game and the win probability and player edge are significantly greater. Thus the Las Vegas game is more desirable from the player's point of view. We may restate the difference between the two games in more practical terms by assuming the monetary value of a chip to be five dollars (the minimum bet allowed in most casinos). Then the player's bets would range from \$5 to \$100 per hand. Table III shows that the Atlantic City player would require initial capital of roughly \$18,110 to participate in the game and that his efforts would be rewarded at an average rate of about 28 cents per hand. By contrast, the Las Vegas player would require roughly \$24,810 initial capital and would expect to win about 52 cents per hand.

Since a human player employing our strategy may expect to receive only about 60 hands per hour in an actual casino, the expected profit in the more favorable Las Vegas game is a modest \$32 per hour on an initial investment of roughly \$24,810. Taking into account the (approximately) 13% chance of actually suffering a loss after twenty-thousand hands of tedious labor, casino blackjack would appear to be a rather unattractive business venture. The case against card-counting is further strengthened by the fact that imperfect shuffling and the presence of other players at the table would result in lower profits than those stated in Table III. In addition, the 1-20 betting ratio employed in the simulation is probably too extreme to be implemented without detection in a casino. The result of detection would likely be immediate shuffle whenever a high bet is made (Atlantic City) or outright expulsion from the game (Las Vegas). One therefore wonders why casinos continually complain about being fleeced by "professional" players.

References

1. R. R. Baldwin, W. E. Cantey, H. Maisel, and J. P. McDermott, The optimum strategy in blackjack, J. Amer. Statist. Assoc., 51 (1956) 429–439.

2. E. O. Thorp, Beat the Dealer, 2nd ed., Vintage, New York, 1966.

3. ______, A favorable strategy for twenty-one, Proc. Nat. Acad. Sci., 47 (1961) 110–112.

4. J. H. Braun, The Development and Analysis of Winning Strategies for the Casino Game of Blackjack, G.B.C. Press, Las Vegas, 1975.

MISCELLANEA

108.

Posthumous Prof?

[The 17th century mathematician] Erhard Weigel [was] a highly respected teacher in his time (and even long after his death).

—Crux Mathematicorum, vol. 8, no. 6.

example, the histogram shows that 40 of the 600 trials resulted in wins of between 1200 and 1400 chips, while only 5 trials resulted in losses of between 1200 and 1400 chips.

The total shaded area is equal to 120,000 chip-occurrences and 87% of this area lies to the right of the origin. A normal curve, normalized to area 120,000, having the same mean and standard deviation as the histogram ($\mu = 2084$, $\sigma = 1786$), is also shown in Fig. 1. The resemblance between histogram and curve supports the hypothesis that the theoretical distribution of player profits is approximately normal. That is, a χ^2 goodness-of-fit test accepts our sample as a random one from the normal distribution shown. Approximate normality of the theoretical distribution is at first somewhat surprising in view of the player's varying bet size and the fact that consecutive hands of blackjack are not strictly independent events. However, a more detailed analysis does lead one to expect approximate normality.

Conclusions. Table III shows that when the same strategy is employed by the player in both the 4-deck and 6-deck games, the expected win is about twice as great in the 4-deck case, while the initial capital required is, roughly, only about one and a half times as great. Furthermore, the relative fluctuation of the player's funds is less in the 4-deck game and the win probability and player edge are significantly greater. Thus the Las Vegas game is more desirable from the player's point of view. We may restate the difference between the two games in more practical terms by assuming the monetary value of a chip to be five dollars (the minimum bet allowed in most casinos). Then the player's bets would range from \$5 to \$100 per hand. Table III shows that the Atlantic City player would require initial capital of roughly \$18,110 to participate in the game and that his efforts would be rewarded at an average rate of about 28 cents per hand. By contrast, the Las Vegas player would require roughly \$24,810 initial capital and would expect to win about 52 cents per hand.

Since a human player employing our strategy may expect to receive only about 60 hands per hour in an actual casino, the expected profit in the more favorable Las Vegas game is a modest \$32 per hour on an initial investment of roughly \$24,810. Taking into account the (approximately) 13% chance of actually suffering a loss after twenty-thousand hands of tedious labor, casino blackjack would appear to be a rather unattractive business venture. The case against card-counting is further strengthened by the fact that imperfect shuffling and the presence of other players at the table would result in lower profits than those stated in Table III. In addition, the 1-20 betting ratio employed in the simulation is probably too extreme to be implemented without detection in a casino. The result of detection would likely be immediate shuffle whenever a high bet is made (Atlantic City) or outright expulsion from the game (Las Vegas). One therefore wonders why casinos continually complain about being fleeced by "professional" players.

References

1. R. R. Baldwin, W. E. Cantey, H. Maisel, and J. P. McDermott, The optimum strategy in blackjack, J. Amer. Statist. Assoc., 51 (1956) 429–439.

2. E. O. Thorp, Beat the Dealer, 2nd ed., Vintage, New York, 1966.

3. ______, A favorable strategy for twenty-one, Proc. Nat. Acad. Sci., 47 (1961) 110–112.

4. J. H. Braun, The Development and Analysis of Winning Strategies for the Casino Game of Blackjack, G.B.C. Press, Las Vegas, 1975.

MISCELLANEA

108.

Posthumous Prof?

[The 17th century mathematician] Erhard Weigel [was] a highly respected teacher in his time (and even long after his death).

—Crux Mathematicorum, vol. 8, no. 6.

They're not ordinary. (See p. 471.)

CONJECTURE 2.

$$f(n) \leqslant \frac{n}{\log n} \text{ for } n \neq 144.$$

Both these conjectures have been verified by computer for $n \leqslant 10{,}000$.

Appendix

n	$f(n)$	n	$f(n)$	n	$f(n)$	n	$f(n)$
1	1	26	2	51	2	76	4
2	1	27	3	52	4	77	2
3	1	28	4	53	1	78	5
4	2	29	1	54	7	79	1
5	1	30	5	55	2	80	12
6	2	31	1	56	7	81	5
7	1	32	7	57	2	82	2
8	3	33	2	58	2	83	1
9	2	34	2	59	1	84	11
10	2	35	2	60	11	85	2
11	1	36	9	61	1	86	2
12	4	37	1	62	2	87	2
13	1	38	2	63	4	88	7
14	2	39	2	64	11	89	1
15	2	40	7	65	2	90	11
16	5	41	1	66	5	91	2
17	1	42	5	67	1	92	4
18	4	43	1	68	4	93	2
19	1	44	4	69	2	94	2
20	4	45	4	70	5	95	2
21	2	46	2	71	1	96	19
22	2	47	1	72	16	97	1
23	1	48	12	73	1	98	4
24	7	49	2	74	2	99	4
25	2	50	4	75	4	100	9

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, Clarendon Press, 1971, p. 239.

2. G. T. Williams, Numbers generated by the function e^{e^x-1}, this MONTHLY, 52 (1945) 323–327.

3. George Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications 2, Gian-Carlo Rota, Editor, Addison-Wesley, Reading, Mass. 1976.

ANSWERS TO PHOTOS ON PAGE 437

No, they are partial. They are two of the most famous partial differential equators in the world. Top: Lars Hörmander of Lund; bottom: Olga Ladyženskaja of Leningrad.

DETERMINING A FAIR BORDER

THEODORE P. HILL
Department of Mathematics, University of Leiden, Wassenaarseweg 80, Postbus 9512, 2300 RA Leiden, The Netherlands

Introduction. Suppose n countries border on a region the ownership of which is in dispute (Fig. 1). Is there a way of partitioning the disputed territory so each country receives a single piece adjacent to itself which it considers at least $1/n$ the total value of the territory, even though different parts of the territory may be valued differently by individual countries? The main purpose of this paper is to show that such fair borders always exist, under the quite natural assumption that each country's value of the territory is nonatomic (i.e., single points have value zero).

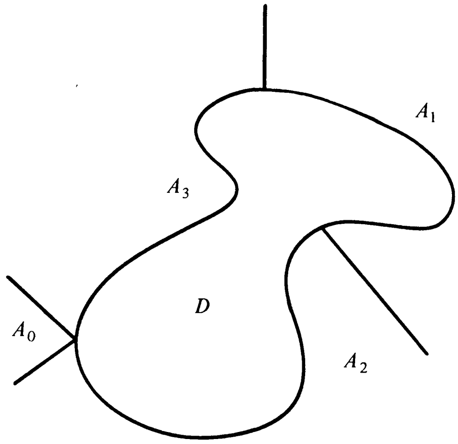


FIG. 1.

This border problem is in some respects closely related to several of the classical fair-division problems such as the "Problem of Cutting a Cake Fairly" of Steinhaus [**2**], [**9**], the "Problem of the Nile" described by Fisher [**4**], [**5**], the "Problem of Similar Regions" of Neyman and Pearson [**8**], and the "Ham Sandwich Problem" of Ulam [**10**].

Steinhaus first raised the question of whether an object (such as a cake) can be divided among n people, who may value different parts of the cake differently, in such a way that each person feels he has received at least $1/n$ the total worth of the cake *according to his own value*. For $n = 2$, the well-known "one cuts, the other chooses" method always yields a solution (although in some respects not an ideal solution, as the second person has an obvious advantage in general). Steinhaus showed that the cake-cutting problem has an affirmative solution for $n = 3$, and then Banach and Knaster solved the problem for general n with the following simple, elegant, and practical solution: pass a long knife parallel to itself slowly over the top of the cake until one of the participants says "stop," cut the cake at that point and give the piece to that participant, and then continue moving the knife. It is easy to see that this procedure guarantees each person at least $1/n$ of the cake, according to his own value, provided that each participant's value of every piece of zero volume is zero.

The author's formal education has included a Bachelor's degree at West Point, a Master's in operations research at Stanford, a year as a Fulbright scholar at Göttingen, and a Ph.D. in mathematics at Berkeley in 1977. He has taught at Washington University and is currently on leave from Georgia Tech as a NATO postdoctoral fellow at the University of Leiden. His main mathematical interests center around stochastic processes (especially abstract gambling theory and optimal stopping theory); this excursion into fair-division problems was largely recreational.

The "Problem of the Nile" [2], [4], [5] involves partitioning a set into k pieces (instead of n, as in the cake-cutting problem) and then evaluating each of n measures on each piece. "Each year the Nile would flood, thereby irrigating or perhaps devastating parts of the agricultural land of a predynastic Egyptian village. The value of the different portions of the land would depend on the height of the flood. In question was the possibility of giving to each of the k residents a piece of land whose value would be $1/k$ of the total land value, no matter what the height of the flood." Feller [3] showed that if there are an infinite number of flood heights possible, the problem need not have a solution, whereas Neyman [7] first proved that a (nonconstructive) solution always exists if there are only a finite number of possible flood heights.

A special case of Neyman and Pearson's "Problem of Similar Regions" [8] which is actually equivalent to the "Problem of the Nile" is the "Bisection Problem": given a set and n probability measures on it, does there always exist a *single* subset having exactly measure one-half with respect to each of the n measures? (More generally, the "Problem of Similar Regions" asks for the existence, for each α in $[0, 1]$, of a single set having exactly measure α with respect to each measure.)

Another classical bisection problem, but one which involves bisection in a particular way (namely by a hyperplane), is that of Ulam's "Ham Sandwich Problem": can an ordinary ham sandwich, consisting of bread, ham, and butter be cut by a plane in such a way that each of the three ingredients is cut exactly in half? More generally, can n objects in euclidean n-space always be simultaneously bisected by a single hyperplane? The answer is affirmative, and the standard proof is an application of the Borsuk-Ulam theorem [1], [2]:

(1) *If f is a continuous map of the sphere in n-dimensional space into $(n - 1)$-dimensional space such that $f(-x) = -f(x)$ for every x, then there is some point on the sphere mapped into the origin.*

(For an interesting discussion, historical background, and proofs of the interrelationships among these classical division problems, the reader is referred to [2].)

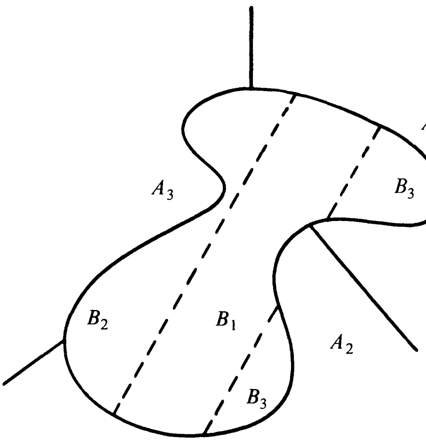


FIG. 2.

Solution of the Border Problem. Solutions of the fair-division problems above are generally not adequate for solution of the border problem for two topological reasons: first, the pieces so determined may be badly disconnected (recall we want each country to receive a *single* piece); and second, the pieces may well not be adjacent to the correct countries (Fig. 2). Moreover, it is not always possible to connect the Banach-Knaster pieces to the correct countries (by thin, nonintersecting roads for example) without destroying either the overall fairness of the division, or the

connectivity of the other pieces. However, a (nonconstructive) cake-cutting result of Dubins and Spanier can be used to help settle the border problem affirmatively. Using Lyapounov's convexity theorem [2], [6],

(2) *The range of every nonatomic, countably additive, finite dimensional, vector-valued measure is convex,*

Dubins and Spanier were able to show there is always a (Borel) partitioning of the cake so that each person receives strictly *more* than his share, provided at least two people value some part of the cake differently.

THEOREM 1 (Dubins and Spanier [2], Corollary 1.2). *Let D be a Borel subset of $\mathbb{R}^k$, and suppose $\mu_1,\dots,\mu_n$ are nonatomic (Borel) probability measures supported on D, with $\mu_i \neq \mu_j$ for some $i \neq j$. Let $p_i > 0$ with $\Sigma^n p_i = 1$. Then there exists a Borel partition $C_1,\dots,C_n$ of D such that $\mu_i(C_i) > p_i$ for each i.*

(The numbers p_i refer to the minimum proportion of the territory to be given to the ith country; often $p_i \equiv 1/n$ for all i.)

It should be noted that the partition guaranteed by Theorem 1 may also fail to solve the border problem for the same reasons as the Banach-Knaster procedure may; moreover, the elements of the Dubins-Spanier partition can be very complicated Borel sets in general.

Countries (and the disputed territory D) will be identified with *open connected regions* in $\mathbb{R}^2$. For a set $A \subset \mathbb{R}^2$, $\bar{A}$ denotes the closure of A, $\mathring{A}$ the interior of A, and ∂A the boundary ($\bar{A}\setminus\mathring{A}$) of A. By a nonatomic measure is meant a Borel measure which assigns measure zero to each singleton set (single point).

Definition. Open connected subsets A and B of $\mathbb{R}^2$ are *adjacent* if $\partial A \cap \partial B$ contains an open arc α (homeomorphic image of $(0, 1)$) such that $A \cup B \cup \alpha$ is open and connected.

In Figure 1, D is adjacent to A_1, A_2 and A_3, but not to A_0. Intuitively, two connected regions are adjacent if they "touch" on an interval, that is, if a single country can be formed from the two regions by erasure of some small open arc on their common boundary. That the common boundary just contain an open arc is not enough to insure this merger can always be accomplished (and hence that a fair division in the "connected-adjacent" sense exists), as can be easily seen by looking at such borders as the $\sin(1/x)$ curve describes. (However, it is easy to check in the above definition that if $A \cup B \cup \alpha$ is open, it is automatically connected.) We are now ready to state the main result; recall that p_i represents the minimum proportion of the disputed territory the ith country is to receive.

THEOREM 2. *Suppose $D, A_1,\dots,A_n$ are open connected regions in $\mathbb{R}^2$ with A_i adjacent to D for each i. If $\mu_1,\dots,\mu_n$ are nonatomic probability measures on D and $p_i \geq 0$, $\Sigma^n p_i = 1$, then there exist disjoint open connected subsets $B_1,\dots,B_n$ of D with B_i adjacent to A_i for all i, $\mu_i(B_i) \geq p_i$ for all i, and with $\overline{\cup B_i} = D$.*

Proof of Theorem 2. Without loss of generality, $p_i > 0$ for all i.

Case 1. $\mu_1 = \cdots = \mu_n$. Expand A_1's territory into D continuously, taking care to avoid the boundary of D and to pass continuously through any arcs of positive μ_j measure, until $\mu_1(\bar{B}_1) = \mu_1(B_1) = p_1$ (Fig. 3). (This is possible since μ_1 is nonatomic and D is connected.) Now $D \setminus \bar{B}_1$ is open and connected, and since $\mu_1 = \cdots = \mu_n$, $\mu_j(D \setminus \bar{B}_1) = 1 - p_1$, for all $j = 1,\dots,n$. Continue similarly to find $B_2,\dots,B_{n-1}$, and let $B_n = D\setminus(\overline{B_1 \cup \cdots \cup B_{n-1}})$. This completes Case 1.

Case 2. $\mu_i \neq \mu_j$ for some $i \neq j$. By Theorem 1 there exists a Borel partition $C_1,\dots,C_n$ of D satisfying $\mu_i(C_i) > p_i$ for all $i = 1,\dots,n$. Since the $\{\mu_i\}$ are Borel, there exist (disjoint) open balls $E_1,\dots,E_k$ in D with $\mu_i(\partial E_j) = 0$ for all i and j, and

$$\mu_i\left(\bigcup^k E_j \Delta C_1\right) < \varepsilon = \min_j\left\{\mu_j\left(C_j\right) - p_j\right\} \text{ for all } i = 1, \ldots, n.$$

Since D is open and connected, it is path (piecewise linear) connected and it follows easily that there are (sausage-shaped) open connected subsets of D adjacent to A_1 the closures of which have arbitrarily small μ_i measure for all i, and which contain all the centers of $E_1, \ldots, E_k$. Let B_1 be the union of one of these sets with $\cup^k E_j$ which satisfies $\mu_i(B_1 \Delta C_1) < \varepsilon$ and $\mu_i(\partial B_1) = 0$ for all $i = 1, \ldots, n$. Then B_1 is open and connected, adjacent to A_1, and satisfies $\mu_1(B_1) > p_1$, and $\mu_j(C_j \setminus \overline{B}_1) > p_j$ for all $j = 2, \ldots, n$. (See B_2 in Fig. 3.) Now $D \setminus \overline{B_1}$ is open and connected, so continue in this manner finding $B_2, \ldots, B_{n-1}$ and then let $B_n = D \setminus \overline{(B_1 \cup \cdots \cup B_{n-1})}$. It is easy to see that the sets $B_1, \ldots, B_n$ satisfy the conclusion of the theorem. □

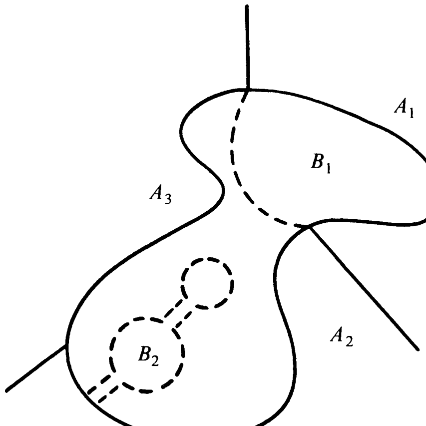


FIG. 3.

If the measures $\mu_1, \ldots, \mu_n$ fail to be nonatomic, the conclusion of the theorem may fail: the worst case is that for some point $x \in D$, $\mu_i(\{x\}) = 1$ for all i.

Theorem 2 may easily be strengthened in two respects. First, the conclusion holds also for $\mathbb{R}^k$ (or k-dimensional manifolds) if one generalizes the definition of adjacency as "touching" on $k - 1$ dimensional regions (i.e., $\partial A \cap \partial B$ contains a homeomorphic image of $(0, 1)^{k-1} \ldots$). And second, the conclusion may also obviously be strengthened to guarantee that the new borders formed are polygonal. If $\mu_i \neq \mu_j$ for some $i \neq j$, the analog of Theorem 2 corresponding to Theorem 1 is also easy to prove: there is always a border which gives each country strictly more than its fair share. As an immediate corollary to Theorem 2, one may drop all adjacency requirements and conclude that a connected cake may be divided so that each person receives a *single* (i.e., connected) piece which is a fair share.

It would perhaps be of interest to find a practical, constructive method for generating the fair border guaranteed by Theorem 2; no such solution is known to the author.

Other Notions of a Fair Share. That each participant receive a piece which he values at least $1/n$ of the total value is certainly not the only criterion for what constitutes a "fair share." Suppose one person receives a piece which he values at exactly $1/n$, whereas others get strictly more than $1/n$ according to his *and* their values. Has the first person received a "fair share"? In the border problem, since D is a nonempty open set, it is easy to see that the solution guaranteed by Theorem 2 is *never* unique, and the question arises of whether or not borders exist which are in some sense optimal, or at least more fair. Dubins and Spanier addressed this question for

cake-cutting, and offered several different criteria for determining whether one partition is better than another.

The first criterion suggested is that the partition $\{A_1, A_2, \ldots, A_n\}$ is better than the partition $\{B_1, \ldots, B_n\}$ if $\Sigma^n \mu_i(A_i) \geqslant \Sigma^n \mu_i(B_i)$, and for cake-cutting they proved [**2**, Theorem 2] that optimal partitions in this sense always exist. On the other hand, for the border problem it may be seen from Fig. 2, with μ_i uniformly distributed on B_i for each i, that optimal partitions in this first sense do not always exist (although clearly "ε-optimal" ones always exist, as can be shown by a slight modification in the proof of Theorem 2 using [**2**, Theorem 2] in place of Theorem 1).

A second notion of optimality Dubins and Spanier suggested was the following: "Find a partition that maximizes the amount received by the person who gets the least, and, among all such partitions, find one which maximizes the amount received by the person who gets next to the least, etc." (for a formal definition, see [**2**, p. 8]). Again, for cake-cutting, optimal partitions in this second sense always exist [**2**, Cor. 6.10] but for the border problem they do not exist in general as can be seen from Fig. 2 again with the uniform distributions given above. However, "ε-optimal" borders in this second sense also always exist; this time use [**2**, Cor. 6.10] in the modification of the proof of Theorem 2.

A third notion of optimal partition suggested in [**2**] is one in which $\mu_i(A_j) = p_j$ for all i and all j, that is, all participants agree that each person received *exactly* the correct amount. Proving an assertion of Steinhaus [**9**], Dubins and Spanier showed that for cake-cutting, optimal partitions in this third sense also always exist [**2**, Cor. 1.1], and again it is easy to use their result to modify the proof of Theorem 2 and conclude that ε-optimal (i.e., $|\mu_i(A_j) - p_j| < \varepsilon$ for all i and j) solutions to the border problem always exist. But whether or not *optimal* borders (in this third sense) always exist is not known to the author, even if D is bounded. The techniques above do not seem to work; simply "taking limits" in general destroys both the connectivity and the correct adjacency of the pieces.

One last criterion for comparing borders which are already fair in one of the above measure-theoretic senses, a criterion suggested by L. Karlowitz, is that of comparing total lengths of borders, and looking for one with minimal length (although some countries with large armies, for example, may well prefer long borders). Another look at Fig. 2 shows that fair borders of minimal length do not exist in general, but if D is bounded, there are always fair borders of nearly minimal length.

Acknowledgements. The author is grateful to George Cain and Victor Pestien for various suggestions, and to the Mathematics Department at the University of Leiden for its hospitality and technical assistance during the academic year 1982-3.

References

1. K. Borsuk, Drei Sätze über die n-dimensionale euklidische Sphäre, Fund. Math., 20 (1933) 177–190.
2. L. Dubins and E. Spanier, How to cut a cake fairly, this MONTHLY, 68 (1961) 1–17.
3. W. Feller, Note on regions similar to the sample space, Statistical Research Memoirs, University of London, 1938, 116–125.
4. R. Fisher, Quelques remarques sur l'estimation en statistique, Biotypologie (1938) 153–159.
5. ______, Uncertain inference, Proc. Amer. Acad. Arts Sci., 71 (1936) 245–257.
6. A. Lyapounov, Sur les fonctions-vecteurs complètement additives, Bull. Acad. Sci. URSS, 4 (1940) 465–478.
7. J. Neyman, Un théorème d'existence, C.R. Acad. Sci. Paris, 222 (1946) 843–845.
8. J. Neyman and E. Pearson, On the problem of the most efficient tests of statistical hypotheses, Philos. Trans. Roy. Soc. London Ser. A, 231 (1933) 289–337.
9. H. Steinhaus, Sur la division pragmatique, Econometrica (supplement), 17 (1949) 315–319.
10. A. Stone and J. Tukey, Generalized sandwich theorems, Duke Math. J., 9 (1942) 356–359.

THE SLOW CONTINUED FRACTION ALGORITHM VIA 2 × 2 INTEGER MATRICES

HARRY APPELGATE AND HIRONORI ONISHI
Department of Mathematics, City College of New York, 138th Street and Convent Avenue, New York, NY 10031

Introduction. It is well known that any real number α may be expanded as a simple continued fraction

$$\alpha = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cdots}},$$

where the numerators are all 1's and q_n, called the partial quotients, are integers with $q_n > 0$ for $n > 0$. The expansion is finite for rational α and infinite for irrational α. Since the simple continued fraction is completely determined by the sequence of partial quotients $q_0, q_1, q_2, \ldots$, it is customary to use the compact notation

$$\alpha = [q_0, q_1, q_2, \ldots].$$

Of considerable interest is the case where the sequence of partial quotients is eventually periodic, e.g.,

$$[1,2,3,5,4,5,4,5,4,\ldots],$$

which is abbreviated as $[1,2,3,\overline{5,4}]$. In 1770, Lagrange proved that the simple continued fraction expansion of a real irrational number is eventually periodic if and only if it is a quadratic irrational, i.e., $a + b\sqrt{\Delta}$, where $a, b \in \mathbb{Q}$ and $\Delta \in \mathbb{N}$ but not a perfect square.

Less well known than Lagrange's theorem are a number of results concerning the fine structure of continued fractions. For example, there is the "palindromic" character of the expansion of $\sqrt{\Delta}$. This is described in (1.8) below, but the idea is well illustrated by

$$\sqrt{19} = [4, \overline{2,1,3,1,2,8}],$$

in which the periodic part contains the palindrome 2, 1, 3, 1, 2 followed by a term 8 which is twice the initial term 4. Theorem 1.8 asserts that this happens for any $\alpha = \sqrt{\Delta}$, $\Delta \in N$, $\Delta \neq n^2$.

Another interesting result is Galois' theorem (1828) characterizing those α that have purely periodic simple continued fractions (cf. (1.7) below).

There are also striking connections with Pell's equation

$$x^2 - \Delta y^2 = \pm 1, \Delta \neq n^2.$$

The equation with $+1$ always has solutions, and the continued fraction of $\sqrt{\Delta}$ shows us how to find all of them. The equation with -1 has solutions if and only if the period of the continued fraction of $\sqrt{\Delta}$ is odd. This is described in (1.9) below.

The purpose of this paper is to present these and related results within a conceptual framework which, we believe, makes the underlying ideas easy to grasp. Our starting point is Richards' paper

Harry Appelgate: I received my Ph.D. in Mathematics from Columbia University in 1965 under Samuel Eilenberg. Before attending graduate school, I worked in computing. Since 1965 I've been teaching at the City College of New York. My main interests are category theory and computing. In recent years I've been working with H. Onishi in integer matrices and algorithms.

Hironori Onishi: I received my Ph.D. in Mathematics from M.I.T. in 1961 under N. C. Ankeny. I taught at the University of Michigan for one year (1961–1962). Since 1963 I have been teaching at the City College of New York. My first interest was in analytic number theory but I have since drifted into algebra and algebraic number theory. In recent years I've been working with H. Appelgate in integer matrices and algorithms.

"Continued fractions without tears" [**4**]. (However, while partially motivated by his paper, our paper is self-contained.) Our main innovation is the use of 2×2 integer matrices to extend Richards' observations, and thus obtain the deeper results outlined above (as well as others).

The fact that there is a connection between 2×2 integer matrices and continued fractions is well known. It arises by considering a fraction x/y as a lattice point (x, y) in the plane. Thus the point (x, y) is identified with the inverse slope of the line through the origin and (x, y). (In technical terms, (x, y) determines a point of the projective line.)

We now proceed to the detailed developments. While no single ingredient in this approach is new, we believe that the "mix" is, which leads to some new observations (cf. (3.13) through (3.18)). Furthermore, we believe that this approach makes the subject conceptual and pretty easy. See if you agree.

Outline. This paper consists of three sections. In § I, we summarize those aspects of the theory of simple continued fractions which we deal with in terms of 2×2 matrices. Seven theorems are stated, whose proofs will come later in §3. In §2 we describe the slowed down version of the simple continued fraction algorithm. Also we lay down the foundation for the use of 2×2 integer matrices and conclude with a clarification of the relationship between the slow algorithm and the "fast algorithm," i.e., the classical simple continued fraction algorithm. In §3 we state and prove theorems which correspond to the classical theorems in §1.

1. Continued Fractions.

(1) The basic step in the simple continued fraction algorithm is given by the function

$$f(\alpha) = \frac{1}{\alpha - q}$$

of a real variable α, where $q = q(\alpha)$ is the largest integer $< \alpha$. (cf. (2) in case you notice a difference.) A real number α determines an infinite sequence $\alpha_0, \alpha_1, \alpha_2, \ldots$ of real numbers such that

$$\alpha_0 = \alpha, \alpha_n = f(\alpha_{n-1}) \quad \text{for } n > 0.$$

α_n is called the nth *complete quotient* of α. The largest integer $q_n < \alpha_n$ is called the nth *partial quotient* of α.

(2) Remark for those who are already acquainted with the traditional treatment of simple continued fraction. Traditionally, q in the basic step is taken to be the largest integer $\leqslant \alpha$, i.e., $q = [\alpha]$. If α is an integer and $q = [\alpha]$, then $q = \alpha$ and $f(\alpha)$ is not defined. Our modification avoids this situation, i.e., $f(\alpha)$ is always defined. With this modified definition of partial quotients q_n, $[q_0, \ldots, q_n]$ should be defined by the recursion

$$[q_0] = q_0 + 1 \text{ and } [q_0, \ldots, q_n] = q_0 + \frac{1}{[q_1, \ldots, q_n]}.$$

(3) The sequence $q_0, q_1, q_2, \ldots$ of partial quotients of a real number α satisfies the following conditions:

(i) $q_n \geqslant 0$ for all $n > 0$ and if $q_k = 0$ for some $k > 0$, then $q_n = 0$ for all $n > k$.

THEOREM. *Given any infinite sequence $q_0, q_1, q_2, \ldots$ of integers satisfying the condition* (i), *there is a unique real number α such that q_n is the nth partial quotient of α for every $n \geqslant 0$.*

(4) Real numbers α and β are said to be *equivalent*, and written $\alpha \sim \beta$, if there are integers a, b, c, d such that

$$\beta = \frac{a\alpha + c}{b\alpha + d} \text{ and } ad - bc = \pm 1.$$

THEOREM. *Given real numbers α and β, $\alpha \sim \beta$ if and only if a complete quotient of α is equal to a complete quotient of β* (cf. [**3**, Theorem 175, p. 142]).

(5) The following theorem is almost trivial.

THEOREM. *A real number α is rational if and only if* 1 *is a complete quotient of α, or equivalently,* 0 *is a partial quotient of α.*

(6) The continued fraction of a real number α is said to be *periodic* if $\alpha_{n+l} = \alpha_n$ for some $n \geqslant 0$ and $l > 0$, where α_n are the complete quotients of α. If $\alpha_{n+l} = \alpha_n$, then $\alpha_{n+kl} = \alpha_n$ for all $k > 0$. In terms of the partial quotients q_n, (the continued fraction of) α is periodic if and only if the sequence $q_n, q_{n+1}, \ldots, q_{n+l-1}$ repeats indefinitely for some $n \geqslant 0$ and $l > 0$. By (5), a rational number is trivially periodic.

LAGRANGE'S THEOREM. *The continued fraction of an irrational real number α is periodic if and only if α is a quadratic irrational* (cf. Theorem 177, p. 144, [**3**]).

(7) The continued fraction of real number α is said to be *purely periodic* if $\alpha_l = \alpha$ for some $l > 0$, i.e., periodic from the very beginning.

GALOIS' THEOREM. *A quadratic irrational α is purely periodic if and only if $\alpha > 1$ and $-1 < \alpha' < 0$, where α' is the conjugate of α* (cf. p. 102, [**2**]).

(8) Let Δ be a positive nonsquare integer. The continued fraction of $\alpha = \sqrt{\Delta}$ is periodic. $\alpha_1 = (\sqrt{\Delta} - q_0)^{-1}$ is purely periodic by (7). Let l be the least positive integer such that $\alpha_{l+1} = \alpha_1$, so that the sequence $q_1, \ldots, q_l$ repeats.

THEOREM (Palindrome Effect). *Under the situation above for $\alpha = \sqrt{\Delta}$,*

$$q_l = 2q_0 \text{ and } q_n = q_{l-n} \text{ for } 0 < n < l.$$

(9) Let Δ be a positive nonsquare integer and consider the Pell equation

$$x^2 - \Delta y^2 = \pm 1,$$

where x and y are integers. Let q_n be the partial quotients of $\alpha = \sqrt{\Delta}$ and define sequences $a_0, a_1, a_2, \ldots$ and $b_0, b_1, b_2, \ldots$ by

$$\begin{aligned} a_n &= q_n a_{n-1} + a_{n-2} && \text{with } a_0 = q_0 \text{ and } a_{-1} = 1, \\ b_n &= q_n b_{n-1} + b_{n-2} && \text{with } b_1 = q_1 \text{ and } b_0 = 1. \end{aligned}$$

THEOREM. *Let l be the least positive integer such that $\alpha_{l+1} = \alpha_1$. Then for every $k > 0$,*

$$a_{kl-1}^2 - \Delta b_{kl-1}^2 = (-1)^{kl}.$$

Moreover, $(x, y) = (a_{kl-1}, b_{kl-1})$ gives all solutions of the Pell equation such that $x > y > 0$. In particular, solutions with -1 occur if and only if l is odd.

2. The Slow Algorithm and 2 × 2 Matrices.

(1) The basic step of the slow algorithm is given by the function

$$g(\alpha) = \begin{cases} \alpha - 1 & \text{if } \alpha > 1 \\ 1 & \text{if } \alpha = 1 \\ \dfrac{\alpha}{1-\alpha} & \text{if } \alpha < 1 \end{cases}$$

of a positive real variable α just as the basic step of the fast algorithm is given by $f(\alpha)$ of (1.1). The significance of this will become clear soon (cf. (8) and (18)).

(2) We are going to use matrices as tools to facilitate computation. We give an explanation of this in the next three paragraphs. (If you already know what it means for an element of the projective linear group $\mathrm{PL}_2(\mathbb{R})$ to act on the projective line $\mathbb{P}$, you may skip these paragraphs and go to (6).) We assume that the reader is familiar with the action of 2×2 real matrices on the vectors in $\mathbb{R}^2$;

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + cy \\ bx + dy \end{pmatrix}.$$

Note that we use column vectors

$$(x, y)^T = \begin{pmatrix} x \\ y \end{pmatrix}.$$

(3) Given a real number α, a vector $(x, y)^T$ such that $x/y = \alpha$ is said to *represent* α. For example, $(\alpha, 1)^T$ represents α. Every vector $(x, y)^T$ such that $y \neq 0$ represents a real number, namely x/y. A vector of the form $(x, 0)^T$ does not represent a real number. To avoid this inconvenience we let it represent "the point at infinity," denoted by ∞, provided $x \neq 0$. Now every nonzero vector in $\mathbb{R}^2$ represents either a real number or ∞. Given a nonzero vector u in $\mathbb{R}^2$, let $[u]$ denote the real number or ∞ represented by u. Note that given nonzero vectors u and v in $\mathbb{R}^2$, $[u] = [v]$ if and only if $u = tv$ for some scalar $t \neq 0$. We sometimes write $u \sim v$ to mean $[u] = [v]$. The union of the set $\mathbb{R}$ of real numbers and the set $\{\infty\}$ will be denoted by $\mathbb{P}$ and we may refer to the elements of $\mathbb{P}$ as *points*. The set $\mathbb{P}$ is in one-to-one correspondence with the set of straight lines through the origin in the plane. (Cf. (5) for a more geometric description of $\mathbb{P}$.)

(4) Let $\mathrm{GL}_2(\mathbb{R})$ denote the group of 2×2 nonsingular real matrices. For any $A \in \mathrm{GL}_2(\mathbb{R})$, $u \in \mathbb{R}^2$ and $t \in \mathbb{R}$,

$$A(tu) = t(Au).$$

Thus given nonzero vectors u and v in $\mathbb{R}^2$, if $u \sim v$, then $Au \sim Av$. Thus for any $A \in \mathrm{GL}_2(\mathbb{R})$ and $\alpha \in \mathbb{P}$, $A\alpha$ makes sense; if α is represented by $u \in \mathbb{R}^2$, then $A\alpha$ is represented Au, i. e.,

$$A[u] = [Au].$$

For example, with $e = (1,1)^T$, $A1 = [Ae]$. If $\alpha \neq \infty$, then since $(\alpha, 1)^T$ represents α,

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\alpha = \frac{a\alpha + c}{b\alpha + d}(= \infty \text{ if } b\alpha + d = 0).$$

On the other hand, since $(1, 0)^T$ represents ∞,

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\infty = \frac{a}{b}(= \infty \text{ if } b = 0).$$

In this way a matrix $A \in \mathrm{GL}_2(\mathbb{R})$ is a map from $\mathbb{P}$ to $\mathbb{P}$. As maps on $\mathbb{P}$, A and tA for any scalar $t \neq 0$ are the same. Since for us the matrices are tools and use them as maps on $\mathbb{P}$, we shall not make any distinction between A and tA for any scalar $t \neq 0$.

(5) Since geometric intuition is going to be very helpful in the sequel, let us look at $\mathbb{P}$ more geometrically. Given a nonzero vector $(x, y)^T$, consider the line through the origin and (x, y) and then consider the point

$$(\cos\theta, \sin\theta) = \left(\frac{\pm x}{\sqrt{x^2 + y^2}}, \frac{\pm y}{\sqrt{x^2 + y^2}}\right)$$

on the unit circle, where $0 \leqslant \theta < \pi$ and hence the signature $\pm$ is such that $\pm y > 0$ or $y = 0$ and $\pm x > 0$. (See Fig. 1.) This gives a representation of $\mathbb{P}$ as the semicircle

$$\{(\cos\theta, \sin\theta) | 0 \leqslant \theta < \pi\}.$$

Under this representation of $\mathbb{P}$, a real number α is represented by the point

$$(\cos\theta, \sin\theta) = \left(\frac{\alpha}{\sqrt{\alpha^2 + 1}}, \frac{1}{\sqrt{\alpha^2 + 1}}\right)$$

and ∞ is represented by $(1, 0)$. This representation of $\mathbb{P}$ as a semicircle is geometrically and intuitively simple but it is not entirely satisfactory because there is an abrupt discontinuity at

$(1,0)$. To remedy this, consider the point

$$(\cos 2\theta, \sin 2\theta)$$

on the unit circle with θ as before. This closes the gap and we now have a representation of $\mathbb{P}$ as the entire unit circle. Since

$$\cos 2\theta = 1 - 2\sin^2\theta \text{ and } \sin 2\theta = 2\cos\theta\sin\theta,$$
$$(\cos 2\theta, \sin 2\theta) = \left(\frac{x^2 - y^2}{x^2 + y^2}, \frac{2xy}{x^2 + y^2}\right).$$

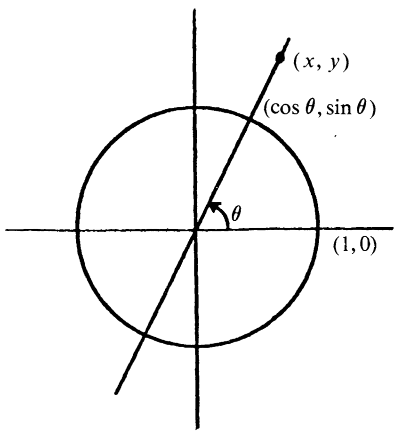


FIG. 1

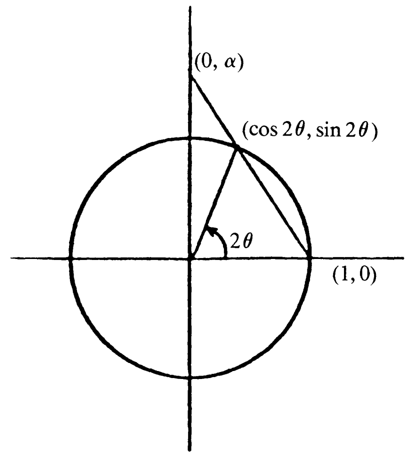


FIG. 2

This representation of $\mathbb{P}$ as the entire unit circle is more wholesome but the intuitive direct relationship between a vector $(x, y)^T$ and the corresponding point is somewhat lost. We shall use this representation in the sequel. The point ∞ is represented by $(1,0)$ and a real number α is represented by

$$\left(\frac{\alpha^2 - 1}{\alpha^2 + 1}, \frac{2\alpha}{\alpha^2 + 1}\right).$$

Since

$$\alpha = \cot\theta = \frac{\cos\theta}{\sin\theta} = \frac{\sin 2\theta}{1 - \cos 2\theta},$$

the point $(\cos 2\theta, \sin 2\theta)$ on the unit circle which represents α is also the intersection of the unit circle with the line through $(0, \alpha)$ and $(1,0)$ as shown in Fig. 2. We shall identify $\mathbb{P}$ with the unit circle under this correspondence. Thus for example, $\infty = (1,0)$ and $3 = (4/5, 3/5)$.

(6) Given $A \in \mathrm{GL}_2(\mathbb{R})$, let (A) denote the set of $\alpha \in \mathbb{P}$ such that $A^{-1}\alpha$ is a positive real number. For example, for the identity I, (I) is the set of positive real numbers in $\mathbb{P}$ or equivalently, (I) is the open arc interval on the unit circle from $(1,0)$ to $(-1,0)$ counter-clockwise. (See Fig. 3.) In general,

$$(A) = \{\alpha \in \mathbb{P} | A^{-1}\alpha \in (I)\}.$$

Since A is one-to-one on $\mathbb{P}$, we also have that

$$(A) = A(I) = \{A\alpha | \alpha \in (I)\}.$$

The set (A) is an open arc interval on the unit circle whose endpoints are represented by the column vectors of A. For example, if

$$A=\begin{pmatrix}3 & -2\\ 1 & -1\end{pmatrix}=(a,b),\, a=\begin{pmatrix}3\\1\end{pmatrix},\, b=\begin{pmatrix}-2\\-1\end{pmatrix},$$

then (A) is the open interval from

$$[a]=\left(\frac{4}{5},\frac{3}{5}\right)=3 \text{ to } [b]=\left(\frac{3}{5},\frac{4}{5}\right)=2$$

clockwise. (See Fig. 4.) Since $A^{-1}\alpha=(\alpha-2)/(\alpha-3)$, in $\mathbb{R}$,

$$(A)\cap\mathbb{R}=\{\alpha\in\mathbb{R}\,|\,\alpha<2 \text{ or } \alpha>3\}.$$

In this example, $\det A<0$. If $\det A>0$, then the interval is from $[a]$ to $[b]$ counter-clockwise, where $A=(a,b)$. It is clear that for any A and $B\in GL_2(\mathbb{R})$,

$$(BA)=B(A).$$

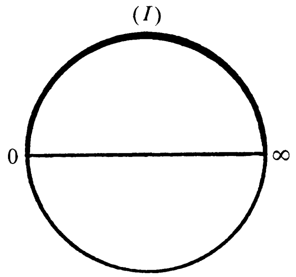


FIG. 3

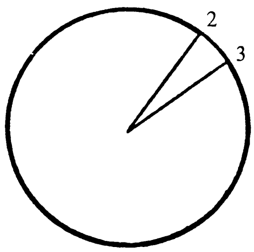


FIG. 4

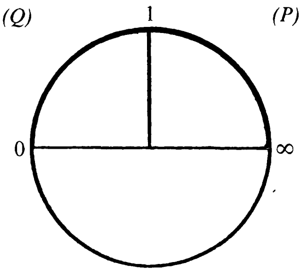


FIG. 5

We let $[A]$ denote the closure of (A), i.e., (A) together with the endpoints of (A).

(7) Put

$$P=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} \text{ and } Q=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}.$$

The intervals (P) and (Q) are shown in Fig. 5. In particular, (I) is the disjoint union of (P), (Q) and 1. For any $A\in GL_2(\mathbb{R})$, if $A=(a,b)$, then

$$AP=(a,Ae) \text{ and } AQ=(Ae,b),$$

where $e=(1,1)^T$. Thus (A) is the disjoint union of (AP), (AQ) and $A1$.

(8) For a real number α,

$$P^{-1}\alpha=\alpha-1 \text{ and } Q^{-1}\alpha=\frac{\alpha}{1-\alpha}.$$

Thus the basic step $g(\alpha)$ of the slow algorithm is given by

$$g(\alpha)=\begin{cases}P^{-1}\alpha & \text{if } \alpha\in(P)\\ 1 & \text{if } \alpha=1\\ Q^{-1}\alpha & \text{if } \alpha\in(Q).\end{cases}$$

(9) We shall refer to the special matrices P and Q as the *letters*. Products of letters are called *words*. We include the identity I as the trivial word. Note that the entries of a word are nonnegative integers and its determinant is $+1$.

(10) Consider a word $L_1\dots L_n$, where L_i are letters. The interval $(L_1\dots L_n)$ is obtained by successive dichotomy of intervals

$$(I) \supset (L_1) \supset (L_1 L_2) \supset \cdots \supset (L_1 \ldots L_n)$$

starting from (I) (cf. (7)). For example, the sequence leading to (PQP) is shown in Fig. 6. Thus we conclude that *every word has a unique spelling*. This observation is fundamental and will be used without explicit reference to it.

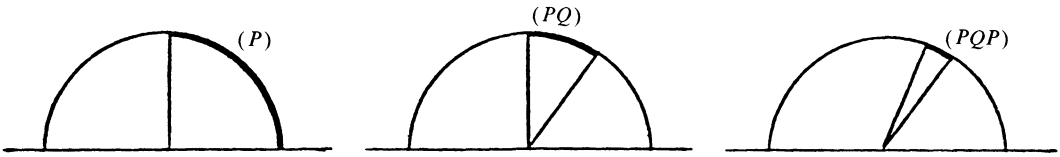


FIG. 6

(11) Put

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

$\det S = -1$, $(S) = (I)$, $S^2 = I$, $SPS = Q$ and $SQS = P$ ($SLS = L^T$ for a letter L). It follows that if X is an n-letter word, then so is SXS. ($SL_1L_2S = SL_1SSL_2S$, etc.) Given α and $\beta \in \mathbb{P}$, we say that α is *congruent* to β, in symbols $\alpha \equiv \beta$, if $\alpha = \beta$ or $S\alpha = 1/\alpha = \beta$. $S\alpha$ is the reflection of α about the line $x = 0$. Given A and $B \in \mathrm{GL}_2(\mathbb{R})$, we say that A is *congruent* to B, in symbols $A \equiv B$, if $A = B$ or $AS = B$. Since $(S) = (I)$, if $A \equiv B$, then $(A) = (B)$.

(12) A rational number or ∞ in $\mathbb{P}$ is represented by a vector having rational components and is referred to as a *rational point* of $\mathbb{P}$. A rational point is always represented by a vector with integer components. A vector with integer components is said to be *primitive* if the components are relatively prime. A rational point is represented by a unique primitive vector $(x, y)^T$ such that $y > 0$ or $y = 0$ and $x > 0$.

(13) Let $\mathrm{GL}_2(\mathbb{Z})$ denote the group of 2×2 integer matrices of determinant ± 1. If $A \in \mathrm{GL}_2(\mathbb{Z})$, then the column vectors of A are primitive. Recall that $A = -A$ as maps on $\mathbb{P}$.

LEMMA. *Given A and $B \in \mathrm{GL}_2(\mathbb{Z})$, $(A) = (B)$ iff $A \equiv B$.*

Proof. If $A \equiv B$, then $(A) = (B)$ as noted in (11). Conversely, suppose that $(A) = (B)$. Put $A = (a, b)$ and $B = (c, d)$. If $\det(AB) = +1$, i.e., $\det A$ and $\det B$ have the same signature, then $[a] = [c]$ and $[b] = [d]$ (cf. (6)). If $\det(AB) = -1$, then $[a] = [d]$ and $[b] = [c]$. Since a, b, c, d are primitive, we get that $A = B$ or $A = BS$. (Recall $A = -A$.)

(14) THEOREM 1. *Given $A \in \mathrm{GL}_2(\mathbb{Z})$, $(A) \subset (I)$ if and only if A is congruent to a word. In particular, if $(A) \subset (I)$ and* $\det A = +1$, *then A is a word.*

Proof. If A is congruent to a word, then the entries of A are $\geqslant 0$ and since $\infty \notin (A)$, $(A) \subset (I)$. Conversely, suppose that $(A) \subset (I)$. Then the entries of A are $\geqslant 0$. From the condition that $\det A = \pm 1$, it is easily verified that if $A = (a, b) \neq I$, then there is no letter L such that $[a] \in (L)$ and $[b] \in (L^T)$. Thus if $A \neq I$, then $(A) \subset (L)$, i. e., $(L^{-1}A) \subset (I)$ for some letter L. The largest entry of $L^{-1}A$ is smaller than that of A. Thus by induction on the size of entries, we conclude that $L^{-1}A$ is congruent to a word. Then A is congruent to a word. The second statement is clear because if X is a word, then $\det X = +1$ and $\det(XS) = -1$.

(15) Given $\alpha \in \mathbb{P}$, it is easy to find $A \in \mathrm{GL}_2(\mathbb{Z})$ such that $\alpha \in (A)$. For example, if $\alpha \in \mathbb{R}$ and k is an integer $< \alpha$, then $\alpha \in (P^k)$. By a *pair* (α, A), we shall mean a pair consisting of $\alpha \in \mathbb{P}$ and $A \in \mathrm{GL}_2(\mathbb{Z})$ such that $\alpha \in (A)$.

(16) A pair (α, A) determines an infinite sequence $\alpha_0, \alpha_1, \alpha_2, \ldots$ of points such that

$$\alpha_0 = A^{-1}\alpha, \quad \alpha_n = L_n^{-1}\alpha_{n-1} \quad \text{for } n > 0,$$

where L_n is P or Q or I according as $\alpha_{n-1} \in (P)$ or $\alpha_{n-1} \in (Q)$ or $\alpha_{n-1} = 1$. We refer to α_n as the nth *tail* of the pair (α, A). The sequence $L_1, L_2, L_3, \ldots$ will be refered to as the *spelling* of the pair (α, A). If X is the n-letter word such that $\alpha \in (AX)$, then

$$\alpha_n = X^{-1}A^{-1}\alpha.$$

(17) If 1 is a tail of (α, A), then clearly α is rational. Conversely:

THEOREM 2. *If α is rational, then for every $A \in \mathrm{GL}_2(\mathbb{Z})$ such that $\alpha \in (A)$, 1 is a tail of (α, A).*

Proof. $\alpha_0 = A^{-1}\alpha$ is rational and is represented by a primitive vector $(a, b)^T$ such that $a > 0$ and $b > 0$. Since

$$P^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x-y \\ y \end{pmatrix} \text{ and } Q^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y-x \end{pmatrix},$$

successive subtraction of the smaller component from the larger gives a word X such that

$$X^{-1}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \text{ i.e., } X^{-1}\alpha_0 = 1.$$

(18) We now clarify the relationship between the slow algorithm and the fast algorithm. Let α be a real number > 1 and let β_n be the nth complete quotient of α to distinguish it from the tail α_n of (α, I). Since $\alpha > 1$, $\alpha \in (P)$ and $\alpha_1 = P^{-1}\alpha$. Let q_0 be the largest integer $< \alpha$. Then q_0 is the largest integer such that $\alpha \in (P^{q_0})$ and

$$\alpha_n = P^{-n}\alpha \quad \text{for } 0 \leqslant n \leqslant q_0.$$

Put $A_1 = P^{q_0}S$. Then

$$A_1^{-1}\alpha = \frac{1}{\alpha - q_0} = \beta_1.$$

Since $\alpha \notin (P^{q_0+1})$, $\alpha_{q_0} \notin (P)$. If $\alpha_{q_0} = 1$, then $\beta_1 = 1$ and $\alpha_n = 1$ for all $n \geqslant q_0$ and $\beta_n = 1$ for all $n \geqslant 1$. If $\alpha_{q_0} \neq 1$, then $\alpha_{q_0} \in (Q)$ and $S\alpha_{q_0} \in (P)$ and $\beta_1 > 1$. Let q_1 be the largest integer $< \beta_1$. Then q_1 is the largest integer such that $S\alpha_{q_0} \in (P^{q_1})$, or equivalently, $\alpha_{q_0} \in (Q^{q_1})$ and

$$\alpha_{q_0+n} = Q^{-n}\alpha_{q_0} \quad \text{for } 0 \leqslant n \leqslant q_1.$$

Put $A_2 = A_1P^{q_1}S = P^{q_0}Q^{q_1}$. Then

$$A_2^{-1}\alpha = \frac{1}{\beta_1 - q_1} = \beta_2.$$

Continuing this, we arrive at the following conclusion: if

$$\underbrace{P, \ldots, P}_{q_0}, \underbrace{Q, \ldots, Q}_{q_1}, \underbrace{P, \ldots, P}_{q_2}, Q, \ldots$$

is the spelling of (α, I), then $q_0, q_1, q_2, \ldots$ is the sequence of partial quotients of α and with $p_n = q_0 + \cdots + q_{n-1}$,

$$S^n\alpha_{p_n} = \beta_n.$$

3. Theorems.

(1) The spelling of a pair (α, A) satisfies the following conditions:

(i) if $L_k = I$ for some $k > 0$, then $L_n = I$ for all $n \geqslant k$, and for any $k > 0$, not all L_n for $n \geqslant k$ are the same letter.

This corresponds to the condition (i) of (1.3).

THEOREM 3. *Given $A \in \mathrm{GL}_2(\mathbb{Z})$ and an infinite sequence $L_1, L_2, L_3, \ldots$, where $L_t = P, Q$ or I, satisfying the condition* (i), *there is a unique $\alpha \in (A)$ such that $L_1, L_2, L_3, \ldots$ is the spelling of (α, A).*

Proof. Put $X_n = L_1 \dots L_n$ and consider the sequence of points $AX_n 1 \in (AX_n)$. Since $(AX_{n+1}) \subset (AX_n)$,

$$\alpha = \lim_{n \to \infty} AX_n 1$$

exists. Clearly, $\alpha \in (A)$ and $\alpha_n = X_n^{-1}A^{-1}\alpha$ is the nth tail of (α, A) for every $n > 0$ and such a point α is unique.

(2) LEMMA. *If $\alpha \in (A) \cap (B)$, where A and $B \in \mathrm{GL}_2(\mathbb{Z})$, then there are words X and Y such that $AX \equiv BY$ and $\alpha \in [BY]$.*

Proof. First suppose that $B = I$. If $\alpha = A1$, then since $\alpha \in (I)$, $(AL) \subset (I)$ for some letter L and α is an endpoint of (AL). If $\alpha \neq A1$, then $\alpha \in (AL)$ for some letter L. Repeating this, we arrive at a word X such that $(AX) \subset (I)$ and $\alpha \in [AX]$. By (1.14), $AX \equiv Y$ for some word Y. In case $B \neq I$, consider the point $B^{-1}\alpha \in (B^{-1}A) \cap (I)$. We have words X and Y such that $B^{-1}AX \equiv Y$ and $B^{-1}\alpha \in [Y]$. Multiplying by B, we get that $AX \equiv BY$ and $\alpha \in [BY]$.

(3) Points α and $\beta \in \mathbb{P}$ are said to be *equivalent*, and written $\alpha \sim \beta$, if $\alpha = C\beta$ for some $C \in \mathrm{GL}_2(\mathbb{Z})$. For α and $\beta \in \mathbb{R}$, this coincides with the relation $\sim$ introduced in (1.4).

(4) Pairs (α, A) and (β, B) are said to be *equivalent*, and written $(\alpha, A) \sim (\beta, B)$, if a tail of (α, A) is congruent to a tail of (β, B). For any pair (α, A) and for any $B \in \mathrm{GL}_2(\mathbb{Z})$,

$$(\alpha, A) \sim (B\alpha, BA).$$

In fact, these pairs have the identical tails.

LEMMA. *The relation $\sim$ on the pairs is an equivalence relation.*

Proof. Only the transitivity needs verification. Suppose that $(\alpha, A) \sim (\beta, B)$ and $(\beta, B) \sim (\gamma, C)$, say $\alpha_k \equiv \beta_l$ and $\beta_m \equiv \gamma_n$. We may assume that $l \leqslant m$. Then $\beta_m = X^{-1}\beta_l$ for some word X. $\gamma_n = S^i\beta_m$ and $\beta_l = S^j\alpha_k$, where i and j are 0 or 1. Thus

$$\gamma_n = S^i X^{-1} S^j \alpha_k.$$

If $j = 0$, then $X^{-1}\alpha_k = \alpha_r$ and $\gamma_n \equiv \alpha_r$ for some r. If $j = 1$, then SXS is a word and $SX^{-1}S\alpha_k = \alpha_r$ and $\gamma_n \equiv \alpha_r$ for some r. Thus $(\alpha, A) \sim (\gamma, C)$.

(5) THEOREM 4. *If $\alpha \in (A) \cap (B)$, where A and $B \in \mathrm{GL}_2(\mathbb{Z})$, then $(\alpha, A) \sim (\alpha, B)$.*

Proof. If α is rational, then 1 is a tail of both (α, A) and (α, B) by (2.17). Assume that α is irrational. By (2), $\alpha \in (AX) = (BY)$ for some words X and Y. $X^{-1}A^{-1}\alpha$ is a tail of (α, A) and $Y^{-1}B^{-1}\alpha$ is a tail of (α, B). Since $AX \equiv BY$, $X^{-1}A^{-1}\alpha \equiv Y^{-1}B^{-1}\alpha$ and $(\alpha, A) \sim (\alpha, B)$.

(6) Given α and $\beta \in \mathbb{P}$, if $(\alpha, A) \sim (\beta, B)$ for some A and $B \in \mathrm{GL}_2(\mathbb{Z})$ such that $\alpha \in (A)$ and $\beta \in (B)$, then clearly $\alpha \sim \beta$. Conversely, if $\alpha \sim \beta$, then $(\alpha, A) \sim (\beta, B)$ for every A and $B \in \mathrm{GL}_2(\mathbb{Z})$ such that $\alpha \in (A)$ and $\beta \in (B)$. In fact, if $\beta = C\alpha$, $C \in \mathrm{GL}_2(\mathbb{Z})$, then by (5),

$$(\alpha, A) \sim (C\alpha, CA) \sim (\beta, B).$$

This result implies the theorem of (1.4).

(7) The theorem of (1.5) follows from (2.17).

(8) A pair (α, A) is said to be *periodic* if $\alpha_{n+l} = \alpha_n$ for some $n \geqslant 0$ and $l > 0$, where α_n are the tails of (α, A). In view of (2.18), it is clear that for $\alpha \in \mathbb{R}$, (α, A) is periodic if and only if the continued fraction of α is periodic.

LEMMA. *If $\alpha_{n+l} = S\alpha_n$, then $\alpha_{n+2l} = \alpha_n$.*

Proof. Let X be the l-letter word such that $\alpha_{n+l} = X^{-1}\alpha_n$. Then $Y = SXS$ is an l-letter word and $\alpha_{n+l} = S\alpha_n \in (SX) = (Y)$. Thus

$$\alpha_{n+2l} = Y^{-1}\alpha_{n+l} = Y^{-1}S\alpha_n = SX^{-1}\alpha_n = \alpha_n.$$

(9) If α is a rational point, then (α, A) is trivially periodic for any $A \in \mathrm{GL}_2(\mathbb{Z})$ such that

$\alpha \in (A)$ (cf. (2.17)). In the rest, we assume that α is irrational; in particular, $\alpha \in \mathbb{R}$. If (α, A) is periodic, then α is a quadratic irrational. In fact, suppose that $\alpha_{n+l} = \alpha_n$, $l > 0$. If X is the word such that $\alpha_{n+l} = X^{-1}\alpha_n$, then $X\alpha_n = \alpha_n$ and α_n is quadratic. Since $\alpha \sim \alpha_n$, α is quadratic also.

(10) LEMMA. *If α is quadratic, then there is $C \in \mathrm{GL}_2(\mathbb{Z})$ such that $C\alpha = \alpha$, $C \neq I$ and* $\det C = +1$.

Proof. A self-contained proof of this lemma will be given later (cf. (20)). For now we give a proof which makes use of a result from algebraic number theory and may be skipped. Let U be the $\mathbb{Z}$-module generated by α and 1 and $K = \mathbb{Q}(\alpha)$ and consider the coefficient ring

$$\mathcal{O}_U = \{x \in K \mid xU \subset U\}.$$

This ring contains a positive unit $\lambda \neq 1$ (cf. Theorem 5, p. 112, [**1**]). Put

$$\lambda\alpha = a\alpha + c,$$
$$\lambda = b\alpha + d,$$

where $a, b, c, d \in \mathbb{Z}$. Then

$$C = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$$

and $C\alpha = \alpha$. Since $\lambda \neq 1$, $C \neq I$. If $\det C = -1$, use C^2; since $\lambda^2 \neq 1$, $C^2 \neq I$.

(11) THEOREM 5. *If α is quadratic, then (α, A) is periodic for any $A \in \mathrm{GL}_2(\mathbb{Z})$ such that* $\alpha \in (A)$ (cf. (1.6)).

Proof. Considering $\alpha_0 = A^{-1}\alpha$, we may assume that $A = I$. By (10), take $C \in \mathrm{GL}_2(\mathbb{Z})$ such that $C\alpha = \alpha$, $C \neq I$ and $\det C = +1$. Since $\alpha = C\alpha \in (C)$, $\alpha \in (I) \cap (C)$. Thus by (2), there are words X and Y such that $\alpha \in (X) = (CY)$. Since $\det C = +1$, $X = CY$. Since $\alpha \in (X)$, $X^{-1}\alpha$ is a tail of (α, I). Since

$$Y^{-1}\alpha = Y^{-1}C^{-1}\alpha = X^{-1}\alpha \in (I),$$

$Y^{-1}\alpha$ is a tail of (α, I) also. If X and Y have the same length, then $X = Y$ and $C = I$. Thus X and Y have different lengths, say one has n letters and the other has $n + l$ letters, $l > 0$. Then $\alpha_{n+l} = \alpha_n$. Thus (α, I) is periodic.

(12) A positive quadratic irrational α is said to be *slowly purely periodic* if (α, I) is purely periodic, i.e., $\alpha_l = \alpha$ for some $l > 0$. For example, for $\alpha = \sqrt{2}$, we have

$$\begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sqrt{2}-1 \\ 1 \end{pmatrix} \xrightarrow{Q} \begin{pmatrix} \sqrt{2}-1 \\ 2-\sqrt{2} \end{pmatrix} \sim \begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix} = S\begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix},$$

where $u \xrightarrow{L} v$ means that $L^{-1}u = v$ for u and $v \in \mathbb{R}^2$. Thus $\alpha_2 = S\alpha$ and $\alpha_4 = \alpha$ and hence $\sqrt{2}$ is slowly purely periodic. If α is slowly purely periodic, then any nontrivial word A such that $A\alpha = \alpha$ is referred to as a *word* for α. For example $PQQP$ is the shortest word for $\sqrt{2}$. A word for α contains both P and Q (cf. (1)). If α is purely periodic (in the fast algorithm), then α is slowly purely periodic. But the converse is false. For example, $\sqrt{2}$ is not purely periodic.

(13) THEOREM 6. *An irrational α is slowly purely periodic if and only if α is a quadratic irrational such that $\alpha > 0$ and $\alpha' < 0$, where α' is the conjugate of α. If α is slowly purely periodic and $A = L_1 \ldots L_l$ is a word for α, then $A^{\mathrm{op}} \overset{\mathrm{df}}{=} L_l \ldots L_1$ is a word for $-\alpha'$.*

Proof. We may assume that $\alpha > 0$. Let L_n be the letter such that $\alpha_n = L_n^{-1}\alpha_{n-1}$. First suppose that α is slowly purely periodic, so that $\alpha_l = \alpha$ for some $l > 0$. Put

$$A = L_1 \ldots L_l = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Then $A\alpha = \alpha$ and we get that

$$b\alpha^2 - (a-d)\alpha - c = 0.$$

Thus α is a quadratic irrational and $\alpha\alpha' = -c/b$. Since $-c/b < 0$, $\alpha' < 0$.

Conversely, suppose that α is a quadratic irrational and $\alpha' < 0$. Let

$$h(t) = at^2 - bt - c$$

be a quadratic polynomial such that $h(\alpha) = 0$, where $a, b, c \in \mathbb{Z}$ and $a > 0$. Since $-c/a = \alpha\alpha' < 0$, $c > 0$. Put

$$C = \begin{pmatrix} b & 2c \\ 2a & -b \end{pmatrix} \text{ and } \Delta = b^2 + 4ac.$$

Then $\det C = -\Delta$ and $C\alpha = \alpha$. Put $C_1 = L_1^{-1}CL_1$. Then $\det C_1 = -\Delta$ and $C_1\alpha_1 = \alpha_1$. We compute that

$$P^{-1}CP = \begin{pmatrix} b-2a & -2h(1) \\ 2a & 2a-b \end{pmatrix} \text{ and } Q^{-1}CQ = \begin{pmatrix} b+2c & 2c \\ 2h(1) & -b-2c \end{pmatrix}.$$

If $L_1 = P$, then $\alpha > 1$ and $h(1) < 0$. If $L_1 = Q$, then $\alpha < 1$ and $h(1) > 0$. Thus the off-diagonal entries of C_1 are > 0 and hence $\alpha_1' < 0$ also. Repeating the same, we get that if

$$C_n = L_n^{-1} \dots L_1^{-1}CL_1 \dots L_n = \begin{pmatrix} b_n & 2c_n \\ 2a_n & -b_n \end{pmatrix},$$

then $C_n\alpha_n = \alpha_n$ and

(1) $$a_n > 0,\ c_n > 0 \text{ and } \det C_n = -\Delta.$$

In particular, $\alpha_n' < 0$. Since there are only a finite number of integer matrices C_n satisfying the condition (1), $C_{n+l} = C_n$ for some $n \geqslant 0$ and $l > 0 (C_0 = C)$. Then α_{n+l} and α_n are the positive root of the same quadratic equation and hence $\alpha_{n+l} = \alpha_n$. Thus $\alpha_{n+l} = \alpha_n$ for some $n \geqslant 0$ and $l > 0$. (If we use (11), this part of the proof is unnecessary. But we wanted a proof which does not depend on the lemma in (10).)

Since $\alpha_n = L_n^{-1}\alpha_{n-1}$, $\alpha_n' = L_n^{-1}\alpha_{n-1}'$. Put

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } \beta_n = J\alpha_n'.$$

Since $\alpha_n' < 0$, $\beta_n > 0$. Since $L_n^TJL_n = J$,

$$L_n^{-T}\beta_n = L_n^{-T}J\alpha_n' = L_n^{-T}JL_n^{-1}\alpha_{n-1}' = J\alpha_{n-1}' = \beta_{n-1}.$$

This says that $\beta_n \in (L_n^T)$ and β_{n-1} is the first tail of (β_n, I). Since $\alpha_{n+l} = \alpha_n$, $\alpha_{n+l}' = \alpha_n'$ and $\beta_{n+l} = \beta_n$. Thus $L_{n+l}^T = L_n^T$ and $L_{n+l} = L_n$ and $\beta_{n+l-1} = \beta_{n-1}$ and hence $\alpha_{n+l-1} = \alpha_{n-1}$. Repeating this, we get that $\alpha_l = \alpha$. Thus α is slowly purely periodic.

We have shown that $L_l^T \dots L_1^T$ is a word for $\beta = \beta_l$. Thus $L_l \dots L_1 = SL_l^T \dots L_1^TS$ is a word for $S\beta = SJ\alpha' = -\alpha'$.

(14) Theorem 6 corresponds to Galois' Theorem of (1.7). The idea of the proof is the same (cf. p. 102, [2]); a slight simplification occurs because the basic step in the slow algorithm is simpler. In any case, the proof does not depend on the lemma in (10). Theorem 6 is the first of the new observations mentioned in the introduction. It brings out a neat, finer structure. Other new observations are consequences of this and are noted in (16), (17) and (18) below.

(15) Let $A \in GL_2(\mathbb{Z})$ and $\det A = +1$. A is a word if and only if A has nonnegative entries (cf. (2.14)). A is a word containing both P and Q if and only if A has positive entries. Given $A \in GL_2(\mathbb{Z})$, put

$$A^* = SAS.$$

If $A = L_1 \dots L_l$ is a word, then since $L_i^* = L_i^T$,

$$A^* = L_1^T \dots L_l^T = (L_l \dots L_1)^T = A^{\mathrm{op}T},$$

and hence

$$A^{\mathrm{op}} = A^{*T} = A^{T*}.$$

$A^{*T} = A$ if and only if AS is symmetric and $A^T = A$ if and only if A is symmetric. Thus for a word A, we have

(i) $A^{\mathrm{op}} = A$ if and only if AS is symmetric,

(ii) $A^{\mathrm{op}} = A^*$ if and only if A is symmetric.

Note that if $A^{\mathrm{op}} = A^*$, then A has an even length. (If A has an odd length, look at the letter in the middle.)

(16) COROLLARY 1. *Let α be a quadratic irrational such that $\alpha > 0$ and $\alpha' < 0$ and let A be a word for α. The word A is a palindrome, i. e., $A^{\mathrm{op}} = A$, if and only if $\alpha' = -\alpha$, i. e., $\alpha = \sqrt{\Delta}$ for some positive nonsquare rational number Δ. If so, then*

$$A = \begin{pmatrix} a & b\Delta \\ b & a \end{pmatrix}$$

for some integers $a > b > 0$.

This is an immediate consequence of Theorem 6. Here are some examples:

Δ	The shortest word for $\sqrt{\Delta}$
2	$PQQP$
3	PQP
5	$PPQQQQPP$
3/5	$QPPPQ$

Corollary 1 gives the theorem of (1.8).

(17) Other interesting conclusions can be drawn from Theorem 6. Let α be a quadratic irrational such that $\alpha > 0$ and $\alpha' < 0$ and let A be a word for α. If $-\alpha' = B^{-1}\alpha$ for some word B shorter than A, then $A = BC$ for some word C and CB is a word for $-\alpha'$ and hence

$$CB = A^{\mathrm{op}} = C^{\mathrm{op}}B^{\mathrm{op}}$$

and both B and C are palindromes. As a special case, if $k = \alpha + \alpha'$ is a positive integer, then $B = P^k$ and $A = P^kC$ and C is a palindrome. For example,

$$A = PPPQQP \quad \text{for } \alpha = 1 + \sqrt{6}.$$

(18) COROLLARY 2. *Let A be a nontrivial word. $A^{\mathrm{op}} = A^*$ if and only if there is a positive nonsquare rational number Δ such that $AS\sqrt{\Delta} = \sqrt{\Delta}$. If so, then*

$$A = \begin{pmatrix} b\Delta & a \\ a & b \end{pmatrix}$$

for some integers $a > b > 0$.

Proof. Put $A = L_1 \ldots L_l$. Given Δ, let $\alpha = \sqrt{\Delta}$ and suppose that $AS\alpha = \alpha$. Then $\alpha_l = A^{-1}\alpha = S\alpha$ and $\alpha_{2l}' = \alpha$. The spelling of $(S\alpha, I)$ is $L_1^T, \ldots, L_l^T, \ldots$. Thus AA^* is a word for α and

$$(AA^*)^{\mathrm{op}} = A^{*\mathrm{op}}A^{\mathrm{op}} = AA^*$$

by (16) and hence $A^{\mathrm{op}} = A^*$. Conversely, suppose that $A^{\mathrm{op}} = A^*$. Then AA^* is a palindrome containing both P and Q. Thus AA^* is a word for $\sqrt{\Delta}$ for some Δ by (16) and $L_1^T, \ldots, L_l^T, \ldots$ is the spelling of (α_l, I). Thus $\alpha_l = A^{-1}\alpha = S\alpha$. The statement about the form of A is clear.

(19) Let Δ be a positive nonsquare rational number and $\alpha = \sqrt{\Delta}$. Let C be the shortest word such that $C^{-1}\alpha \equiv \alpha$, i.e., $CS^i\alpha = \alpha$, $i = 0$ or 1. If $C\alpha = \alpha$, then there is no word A such that $AS\alpha = \alpha$ and every word for α is a power of C. Suppose that $CS\alpha = \alpha$. Then CC^* is the shortest word for α and every word for α is a power of CC^* and every word A such that $AS\alpha = \alpha$ is of the form $(CC^*)^kC$ for some $k > 0$.

If A is a nontrivial word such that $AS^i\alpha = \alpha$, then by (16) and (18),

$$AS^i = \begin{pmatrix} x & \Delta y \\ y & x \end{pmatrix}$$

for some integers $x > y > 0$ and

$$x^2 - \Delta y^2 = \det(AS^i) = (-1)^i.$$

Conversely, let $x > y > 0$ be integers such that

$$x^2 - \Delta y^2 = (-1)^i.$$

Put

$$A = \begin{pmatrix} x & \Delta y \\ y & x \end{pmatrix} S^i.$$

Then $AS^i\alpha = \alpha$ and $\det A = +1$. Thus by (16) and (18),

$$A = C^k \text{ or } (CC^*)^k \text{ or } (CC^*)^k C.$$

This tells us how to find all solutions of the Pell equation (cf. (1.9)).

(20) Finally, we present a self-contained proof of the lemma in (10). Let α be a quadratic irrational and

$$h(t) = at^2 - bt - c$$

be a quadratic polynomial such that $h(\alpha) = 0$, where $a, b, c \in \mathbb{Z}$ and $a > 0$. Its discriminant $\Delta = b^2 + 4ac$ is a positive nonsquare integer. Put

$$C = \begin{pmatrix} b & 2b \\ 2a & -c \end{pmatrix}.$$

Then $C\alpha = \alpha$. Take positive integers x and y such that $x^2 - \Delta y^2 = 1$ by (19) and put

$$A = xI + yC.$$

Then $A\alpha = \alpha$ and $\det A = x^2 - \Delta y^2 = 1$. Clearly $A \neq I$. Thus A is a desired matrix in $\mathrm{GL}_2(\mathbb{Z})$.

References

1. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.
2. Harold Davenport, The Higher Arithmetic, Harper Torchbooks, Harper and Row, New York, 1960.
3. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 3rd ed., Oxford Univ. Press, New York, 1954.
4. Ian Richards, Continued fractions without tears, Math. Mag., 54 (4) (1981) 163–171.

MISCELLANEA

109. **The Good Old Days—Half a Century Ago**

The preparation of the manuscript has been carried on by correspondence without expense to the National Research Council and without special grant for relief from teaching from any of the institutions represented. It is hoped that the time saved to readers by the exposition and references here contained may justify the time that was sacrificed to the preparation of this report.

THE COMMITTEE

—Albert A. Bennett, William E. Milne, and Harry Bateman, Numerical integration of differential equations, *Bulletin of the National Research Council*, 92 (1933).

If A is a nontrivial word such that $AS^i\alpha = \alpha$, then by (16) and (18),

$$AS^i = \begin{pmatrix} x & \Delta y \\ y & x \end{pmatrix}$$

for some integers $x > y > 0$ and

$$x^2 - \Delta y^2 = \det(AS^i) = (-1)^i.$$

Conversely, let $x > y > 0$ be integers such that

$$x^2 - \Delta y^2 = (-1)^i.$$

Put

$$A = \begin{pmatrix} x & \Delta y \\ y & x \end{pmatrix} S^i.$$

Then $AS^i\alpha = \alpha$ and $\det A = +1$. Thus by (16) and (18),

$$A = C^k \text{ or } (CC^*)^k \text{ or } (CC^*)^k C.$$

This tells us how to find all solutions of the Pell equation (cf. (1.9)).

(20) Finally, we present a self-contained proof of the lemma in (10). Let α be a quadratic irrational and

$$h(t) = at^2 - bt - c$$

be a quadratic polynomial such that $h(\alpha) = 0$, where $a, b, c \in \mathbb{Z}$ and $a > 0$. Its discriminant $\Delta = b^2 + 4ac$ is a positive nonsquare integer. Put

$$C = \begin{pmatrix} b & 2b \\ 2a & -c \end{pmatrix}.$$

Then $C\alpha = \alpha$. Take positive integers x and y such that $x^2 - \Delta y^2 = 1$ by (19) and put

$$A = xI + yC.$$

Then $A\alpha = \alpha$ and $\det A = x^2 - \Delta y^2 = 1$. Clearly $A \neq I$. Thus A is a desired matrix in $\mathrm{GL}_2(\mathbb{Z})$.

References

1. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.
2. Harold Davenport, The Higher Arithmetic, Harper Torchbooks, Harper and Row, New York, 1960.
3. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 3rd ed., Oxford Univ. Press, New York, 1954.
4. Ian Richards, Continued fractions without tears, Math. Mag., 54 (4) (1981) 163–171.

MISCELLANEA

109. **The Good Old Days—Half a Century Ago**

The preparation of the manuscript has been carried on by correspondence without expense to the National Research Council and without special grant for relief from teaching from any of the institutions represented. It is hoped that the time saved to readers by the exposition and references here contained may justify the time that was sacrificed to the preparation of this report.

THE COMMITTEE

—Albert A. Bennett, William E. Milne, and Harry Bateman, Numerical integration of differential equations, *Bulletin of the National Research Council*, 92 (1933).

AN INTERESTING CANTOR SET

W. A. COPPEL

Department of Mathematics, IAS, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, Australia

1. Introduction. Cantor sets are all the same topologically, since any two are homeomorphic. But the classical Cantor set also enjoys a number of properties which are not preserved under homeomorphism. These properties extend in part to the Cantor set described in this paper. I was led to its definition by a study of recent work of Feigenbaum [**2**], [**3**]. However, it is of interest in its own right and will be presented from this point of view here. The discussion illustrates a variety of significant topics, such as the notion of dimension, Fourier–Stieltjes transforms, 2-adic integers, and ergodic theory.

2. The Two-Ratio Cantor Set. We define a *Cantor set* C to be a compact metric space which has no isolated point and which has the property that for any two distinct points a, b, there exist disjoint closed sets A, B with union C containing a, b respectively. A compact subset of the real line is a Cantor set if and only if it has no isolated point and contains no interval.

Let r_1, r_2 be positive numbers with sum less than 1. We construct a Cantor set $C = C(r_1, r_2)$ in the following way. Put

$$\begin{aligned} E_0 &= [0,1], \\ E_1 &= [0, r_1] \cup [1 - r_2, 1], \\ E_2 &= [0, r_1^2] \cup [r_1(1 - r_2), r_1] \\ &\quad \cup [1 - r_2, 1 - r_2(1 - r_2)] \cup [1 - r_1 r_2, 1], \\ &\quad \ldots . \end{aligned}$$

In general, E_k is a union of 2^k disjoint compact intervals. If $[a, b]$ is a typical interval in E_k, the intervals in E_{k+1} are given by $[r_1 a, r_1 b]$ and $[1 - r_2 b, 1 - r_2 a]$. It should be observed that if we choose c so that $r_1 < c < 1 - r_2$, then all intervals $[r_1 a, r_1 b]$ lie to the left of c and all intervals $[1 - r_2 b, 1 - r_2 a]$ lie to the right of c. By induction we easily see that $E_{k+1} \subset E_k$ and that E_k has Lebesgue measure $(r_1 + r_2)^k$. More precisely, if the lengths of the $m = 2^k$ intervals in E_k are, from left to right, $d_1, \ldots, d_m$, then the lengths of the $2m$ intervals in E_{k+1}, also from left to right, are

$$r_1 d_1, r_2 d_1, r_2 d_2, r_1 d_2, \ldots, r_1 d_{m-1}, r_2 d_{m-1}, r_2 d_m, r_1 d_m.$$

Thus if we set

$$C = \bigcap_{k=0}^{\infty} E_k,$$

then C is a nonempty compact set of Lebesgue measure zero. A point belongs to C if and only if it is either an endpoint of an interval of some set E_k or a limit of such endpoints. It follows readily that C is a Cantor set.

For $r_1 = r_2 = \frac{1}{3}$ we obtain the original set of Cantor himself. The more general case $r_1 = r_2 = r$ is of some interest in harmonic analysis and is discussed, for example, in Kahane and Salem [**6**].

We note first that *the set $C = C(r_1, r_2)$ has Hausdorff dimension α, where α is the unique root between* 0 *and* 1 *of the equation*

$$r_1^\alpha + r_2^\alpha = 1. \tag{1}$$

The author is a Professorial Fellow at the Institute of Advanced Studies, Australian National University, Canberra. His only degree, B.A., was obtained from Melbourne University in 1950. He has published research on various aspects of the theory of ordinary differential equations, iteration, and mathematical control theory. His article on J. B. Fourier in this MONTHLY, 76 (1969) 468–483, received a 1970 Lester R. Ford award.

The definition of Hausdorff measure is given by Kahane and Salem, and a proof may be modelled on one given there for the case $r_1 = r_2 = r$. In that case $\alpha = \log 2/\log r^{-1}$ can be expressed in closed form. The result is also contained in Theorem II of Moran [7].

There are other reasonable definitions of dimension besides that of Hausdorff. We show next that one proposed by Besicovitch and Taylor [**1**] gives the same value in the present case, although in general the two definitions do not agree.

In forming E_1 we exclude from E_0 an open interval of length τ. In forming E_2 we exclude from E_1 two open intervals of lengths $r_1\tau$ and $r_2\tau$. In general, in forming E_{k+1} we exclude from E_k 2^k open intervals of which one has length $r_1^k\tau$, a further k have length $r_1^{k-1}r_2\tau$, another $k(k-1)/2$ have length $r_1^{k-2}r_2^2\tau,\dots,$ and one has length $r_2^k\tau$. For any β such that $0 < \beta \leqslant 1$ the sum of the βth powers of the lengths of these intervals is

$$\tau^\beta \sum_{j=0}^{k} \binom{k}{j} r_1^{(k-j)\beta} r_2^{j\beta} = \tau^\beta \eta^k, \tag{2}$$

where we have put $\eta = r_1^\beta + r_2^\beta$. Thus the sum of the βth powers of the lengths of all excluded intervals is

$$\sigma_\beta = \tau^\beta(1 + \eta + \eta^2 + \cdots).$$

Hence

$$\begin{aligned} \sigma_\beta &= \tau^\beta(1-\eta)^{-1} \text{ if } \eta < 1, \\ &= \infty \text{ if } \eta \geqslant 1. \end{aligned} \tag{3}$$

The Besicovitch–Taylor dimension is the infimum α of all β for which σ_β is finite and thus satisfies $r_1^\alpha + r_2^\alpha = 1$.

Since the Hausdorff and Besicovitch–Taylor dimensions are equal, it follows by a result of Hawkes [4] that the entropy dimension of C also exists and has the same value α. The entropy dimension is defined by dividing $[0,1]$ into n equal subintervals, counting the number c_n of subintervals which contain points of C, and taking the limit of $\log c_n/\log n$ as $n \to \infty$.

3. The Corresponding Cantor Function. We are now going to construct a distribution function ψ with support on the Cantor set C. Let $I_{k,j}(j = 1,\dots,2^k - 1)$ denote the open intervals complementary to E_k, numbered from left to right. Let ψ_k be the unique continuous function on $E_0 = [0,1]$ such that $\psi_k(0) = 0$, $\psi_k(1) = 1$, ψ_k is linear on each interval of E_k, and

$$\psi_k(x) = j2^{-k} \text{ if } x \in I_{k,j}(j = 1,\dots,2^k - 1).$$

Then ψ_k is nondecreasing, $\psi_{k+1} = \psi_k$ on $I_{k,j}(j = 1,\dots,2^k - 1)$, and

$$|\psi_{k+1}(x) - \psi_k(x)| < 2^{-k}$$

for all $x \in E_0$. Hence the sequence $\{\psi_k\}$ converges uniformly on E_0. Its limit ψ is a continuous, nondecreasing function such that $\psi(0) = 0$, $\psi(1) = 1$, and ψ is constant on any open interval complementary to any set E_k.

This construction can be applied to an arbitrary Cantor set on the real line. However, the special structure of the set C makes it possible to say more. In fact ψ satisfies the functional equations

$$\begin{aligned} \psi(r_1x) &= \frac{1}{2}\psi(x), \\ \psi(1 - r_2x) &= 1 - \frac{1}{2}\psi(x), \quad (0 \leqslant x \leqslant 1) \\ \psi[(1-x)r_1 + x(1-r_2)] &= \frac{1}{2}. \end{aligned} \tag{4}$$

The first two relations are easily verified if x is an endpoint of an interval $I_{k,j}$ and extend at once to arbitrary $x \in E_0$.

The functional equations (4) completely characterize ψ. In fact if ψ^* is any bounded function which satisfies (4), then the difference $\phi = \psi - \psi^*$ satisfies

$$\begin{aligned}\phi(x) &= \frac{1}{2}\phi(x/r_1) \quad (0 \leqslant x \leqslant r_1),\\ &= -\frac{1}{2}\phi[(1-x)/r_2] \quad (1-r_2 \leqslant x \leqslant 1),\\ &= 0 \quad (r_1 \leqslant x \leqslant 1-r_2).\end{aligned}$$

If we denote by μ, μ_1, μ_2 the supremum of $|\phi(x)|$ over the intervals $[0,1], [0,r_1], [1-r_2, 1]$ respectively, then $\mu = \max(\mu_1, \mu_2)$. On the other hand $\mu_1 = \frac{1}{2}\mu$, $\mu_2 = \frac{1}{2}\mu$. It follows that $\mu = 0$.

There is no difficulty in principle in determining the moments of ψ. It is readily shown that

$$\int_0^1 x\, d\psi_k(x) = \frac{1}{2}(1 + h + \cdots + h^k),$$

where $h = (r_1 - r_2)/2$. Letting $k \to \infty$ we obtain

$$\int_0^1 x\, d\psi(x) = \frac{1}{2}(1-h)^{-1}. \tag{5}$$

It may further be shown that

$$\left(2 - r_1^2 - r_2^2\right)\int_0^1 x^2\, d\psi(x) = 1 - r_2(1-h)^{-1}. \tag{6}$$

It is also of interest to consider the Fourier–Stieltjes transform

$$\hat{\psi}(\lambda) = \int_0^1 e^{-i\lambda x}\, d\psi(x) \quad (-\infty < \lambda < \infty).$$

We have

$$\begin{aligned}\hat{\psi}(r_1\lambda) &= \int_0^1 e^{-i\lambda r_1 x}\, d\psi(x)\\ &= 2\int_0^1 e^{-i\lambda r_1 x}\, d\psi(r_1 x)\\ &= 2\int_0^{r_1} e^{-i\lambda x}\, d\psi(x)\end{aligned}$$

and similarly

$$e^{-i\lambda}\hat{\psi}(-r_2\lambda) = 2\int_{1-r_2}^1 e^{-i\lambda x}\, d\psi(x).$$

It follows that $\hat{\psi}$ satisfies the functional equation

$$\hat{\psi}(r_1\lambda) + e^{-i\lambda}\hat{\psi}(-r_2\lambda) = 2\hat{\psi}(\lambda). \tag{7}$$

Harmonic analysts may be interested to determine for what values of r_1 and r_2 the set C is a set of uniqueness.

4. Connection with the 2-adic Integers. We define a 2-adic integer to be an infinite sequence $\alpha = (a_0, a_1, a_2, \ldots)$, where $a_i = 0$ or 1 for all i. If $\beta = (b_0, b_1, b_2, \ldots)$ is another such sequence the sum

$$\alpha + \beta = (c_0, c_1, c_2, \ldots)$$

is defined in the following way. If $a_0 + b_0 < 2$, then $c_0 = a_0 + b_0$, but if $a_0 + b_0 \geqslant 2$, then $c_0 = a_0 + b_0 - 2$ and we carry 1 to the next position. The terms $c_1, c_2, \ldots$ are successively

determined in the same fashion. With this definition of addition the set J of all 2-adic integers is an abelian group.

We can also define a metric on J by setting $d(\alpha, \alpha) = 0$ and $d(\alpha, \beta) = 2^{-k}$ if $\alpha \neq \beta$ and k is the least integer such that $a_k \neq b_k$. This metric is invariant and nonarchimedean, i.e., for all $\alpha, \beta, \gamma \in J$

$$d(\alpha + \gamma, \beta + \gamma) = d(\alpha, \beta),$$
$$d(\alpha + \beta, 0) \leqslant \max[d(\alpha, 0), d(\beta, 0)].$$

Moreover J is now a compact topological group.

If $\delta = (1,0,0,\dots)$, then the multiples $n\delta (n = 0,1,2,\dots)$ consist precisely of all $\alpha = (a_0, a_1, a_2, \dots)$ with $a_i = 0$ for all large i. Hence the semigroup J_0 formed by these multiples is dense in J.

Now let u and v denote the maps of the unit interval into itself defined by

$$u(x) = r_1 x, v(x) = 1 - r_2 x \quad (0 \leqslant x \leqslant 1).$$

Then every endpoint, other than 0 and 1, of an interval of the set E_k can be uniquely represented in the form

$$w_m \circ \cdots \circ w_1(1),$$

where $w_i = u$ or v for each i and $1 \leqslant m \leqslant k$. For example, $v \circ v \circ u(1)$ represents the endpoint $1 - r_2(1 - r_1 r_2)$ of E_3. To such an endpoint we make correspond the 2-adic integer

$$\alpha = (a_0, a_1, a_2, \dots),$$

where $a_i = 0$ for $i > m$, $a_m = 1$, and $a_i = 0$ or 1 according as $w_{m-i} = u$ or v for $0 \leqslant i < m$. To the endpoints 1 and 0 we make correspond the 2-adic integers $\delta = (1,0,0,\dots)$ and $0 = (0,0,0,\dots)$. In this way we define a 1-1 map ω of the set C_0 of all endpoints of intervals of the sets E_k onto the set J_0 of all 2-adic integers $\alpha = (a_0, a_1, a_2, \dots)$ with $a_i = 0$ for all large i.

We will show that this map ω is uniformly continuous. If D is an interval of the set E_k, then it has the form

$$D = w_k \circ \cdots \circ w_1 E_0,$$

where $w_1, \dots, w_k$ are uniquely determined u's or v's. One endpoint of D is $\xi = w_k \circ \cdots \circ w_1(1)$. If $w_i = u$ for $1 \leqslant i \leqslant k$, then the other endpoint of D is $\bar{\xi} = 0$. Otherwise there exists an h $(1 \leqslant h \leqslant k)$ such that $w_h = v$ and $w_i = u$ for all $i < h$, and the other endpoint of D is then

$$\bar{\xi} = w_k \circ \cdots \circ w_{h+1}(1).$$

If $\alpha = \omega(\xi)$ and $\bar{\alpha} = \omega(\bar{\xi})$, then in any event we have $d(\alpha, \bar{\alpha}) \leqslant 2^{-k}$. Any point ξ' of C_0 in the interior of D has the form

$$\xi' = w'_l \circ \cdots \circ w'_1(1),$$

where $w'_1, \dots, w'_l$ are uniquely determined u's or v's and $l > k$. Moreover

$$D' = w'_l \circ \cdots \circ w'_1 E_0 \subset D = w_k \circ \cdots \circ w_1 E_0.$$

Since also

$$D' \subset w'_l \circ \cdots \circ w'_{l-k+1} E_0$$

we must have $w'_{l-i} = w_{k-i}$ for $0 \leqslant i < k$. If $\alpha' = \omega(\xi')$, it follows that $d(\alpha, \alpha') \leqslant 2^{-k}$.

Put

$$r_0 = \min(r_1, r_2, 1 - r_1 - r_2).$$

Then the distance between any two distinct endpoints of intervals of E_k is at least r_0^k. Thus if two distinct points ξ_1, ξ_2 of C_0 are distant less than r_0^k, then they lie in the same interval D of E_k. Hence, by what we have just shown, the corresponding 2-adic integers α_1, α_2 satisfy $d(\alpha_1, \alpha_2) \leqslant 2^{-k}$.

This proves that ω is uniformly continuous. The inverse map ω^{-1} is also uniformly continuous. For if $d(\alpha_1, \alpha_2) \leqslant 2^{-k}$, then $\xi_1 = \omega^{-1}(\alpha_1)$ and $\xi_2 = \omega^{-1}(\alpha_2)$ lie in a common interval $D = w_k \circ \cdots \circ w_1 E_0$ and hence are distant at most r^k, where $r = \max(r_1, r_2)$.

It follows that the map ω admits a unique continuous extension, which we will still denote by ω, mapping the whole of C into J. Moreover ω must map C onto J, since its range is compact and contains a dense subset of J. Since the inverse map also admits a unique continuous extension, ω is actually a homeomorphism of C onto J. If we now define the sum $\xi = \xi_1 \oplus \xi_2$ of two elements of C by $\omega(\xi) = \omega(\xi_1) + \omega(\xi_2)$, then C acquires the structure of a compact topological abelian group. Moreover this structure is naturally connected to the definition of the set C.

Conversely, the measure ν on C determined by the distribution function ψ can be transferred to a measure μ on J. In fact this is precisely the Haar measure on J. This may be shown without difficulty from the explicit form for Haar measure on J, given in Hewitt and Ross [**5**, p. 202]. Returning to C, we see that the measure ν is invariant under the group action just defined. If we set $T\xi = \xi \oplus 1$, for any $\xi \in C$, then it follows from the ergodic theory of group rotations, described in Walters [**9**, pp. 160–162], that for any continuous function $f: C \to \mathbb{R}$

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\xi) \to \int_C f(x)\, d\psi(x) \text{ uniformly as } n \to \infty.$$

It may be noted that the measure ν is also invariant under the piecewise linear transformation S of the unit interval defined by

$$\begin{aligned} Sx &= r_1^{-1}x \quad \text{for } 0 < x < r_1, \\ &= (1 - r_1 - r_2)^{-1}(x - r_1) \quad \text{for } r_1 < x < 1 - r_2, \\ &= r_2^{-1}(1 - x) \quad \text{for } 1 - r_2 < x < 1. \end{aligned}$$

In fact the inverse image of $(0, x)$ is the union of the intervals

$$(0, r_1x), \quad (r_1, r_1 + (1 - r_1 - r_2)x), \quad (1 - r_2x, 1),$$

whose total ν-measure is

$$\psi(r_1x) + 1 - \psi(1 - r_2x) = \psi(x).$$

Transformations of this type have been extensively studied in ergodic theory; see, e.g., Parry [**8**] and Wilkinson [**10**].

References

1. A. S. Besicovitch and S. J. Taylor, On the complementary intervals of a linear closed set of zero Lebesgue measure, J. London Math. Soc., 29 (1954) 449–459.

2. M. J. Feigenbaum, Quantitative universality for a class of nonlinear transformations, J. Statist. Phys., 19 (1978) 25–52.

3. M. J. Feigenbaum, The universal metric properties of nonlinear transformations, J. Statist. Phys., 21 (1979) 669–706.

4. J. Hawkes, Hausdorff measure, entropy, and the independence of small sets, Proc. London Math. Soc., (3) 28 (1974) 700–724.

5. E. Hewitt and K. A. Ross, Abstract Harmonic Analysis, vol. 1, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1963.

6. J.-P. Kahane and R. Salem, Ensembles parfaits et séries trigonométriques, Actualités scientifiques et industrielles 1301, Hermann, Paris, 1963.

7. P. A. P. Moran, Additive functions of intervals and Hausdorff measures, Proc. Cambridge Philos. Soc., 42 (1946) 15–23.

8. W. Parry, Symbolic dynamics and transformations of the unit interval, Trans. Amer. Math. Soc., 122 (1966) 368–378.

9. P. Walters, An Introduction to Ergodic Theory, Graduate Texts in Mathematics 79, Springer-Verlag, Berlin-Heidelberg-New York, 1982.

10. K. M. Wilkinson, Ergodic properties of a class of piecewise linear transformations, Z. Wahrsch. Verw. Gebiete, 31 (1975) 303–328.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N IN4.*

HOW FEW n-PERMUTATIONS CONTAIN ALL POSSIBLE k-PERMUTATIONS?

PETER J. SLATER
Mathematics Department, University of Alabama, Huntsville, AL 35899

If $S_n = \{1, 2, \ldots, n\}$ and $\sigma\colon S_n \to S_n$ is a permutation, then we can represent σ by $(s_1, s_2, \ldots, s_n)$ where $s_i = \sigma(i)$ for $1 \leqslant i \leqslant n$. The collection

$$C = \{(1,2,3,4,5,6), (6,5,4,3,2,1), (3,6,1,5,4,2), (5,4,6,2,1,3), (2,1,5,6,3,4), (4,3,2,1,6,5)\}$$

of permutations of S_6 has the interesting property that for any 2-element subset $\{a, b\}$ of S_6 there are permutations τ' and τ'' in C with $\tau'(a) = b$ and $\tau'(b) = a$ and with $\tau''(a) = a$ and $\tau''(b) = b$. In other words, any permutation of a 2-element subset of S_6 is "contained in" a member of C.

Problem. Determine $M(n, k)$, the minimum number m of permutations $\sigma_1, \sigma_2, \ldots, \sigma_m$ of S_n such that, for any k-element subset K of S_n and any permutation τ of K, at least one σ_i satisfies $\sigma_i(j) = \tau(j)$ for all $j \in K$.

Trivially, $M(n, n) = n!$ and $M(n, 1) = 1$, and it is easy to see that $M(2p - 1, 2) = M(2p, 2) = 2p$. However, the computation of $M(n, 3)$ is already quite complex. The values for $n = 3$, 4, 5 and 6 are 6, 15, 30 and 35.

Consider $M(n, n - 1)$. If $K = S_n - h$ and τ is a permutation of K, and if $\sigma_i(j) = \tau(j)$ for $j \in K$, then $\sigma_i(h) = h$ and σ_i has a fixed point. In short, a permutation σ of S_n is the unique extension of at least one permutation on $n - 1$ elements of S_n if and only if σ has at least one fixed point. As in Brualdi [**1**, page 79] a permutation with no fixed points is called a derangement, and d_n denotes the number of derangements of S_n. Then

$$\begin{aligned} M(n, n-1) &= n! - d_n \\ &= n! - n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right), \end{aligned}$$

and the resulting sequence [**2**, #1423]

$$1, 1, 4, 15, 76, 455, 3186, 25487, \ldots$$

first appeared in a study of the game of Mousetrap [**3**].

It is not true in general that $M(n, n - h)$ is the number of permutations of S_n with at least h points fixed. For example, there are six permutations of S_4 with exactly two fixed points and one with exactly four, but $M(4, 2) = 4$.

References

1. R. A. Brualdi, Introductory Combinatorics, North Holland, New York, 1977.
2. N. J. A. Sloane, A Handbook of Integer Sequences, Academic Press, New York, 1973.
3. A. Steen, Some formulae respecting the game of Mousetrap, Quart. J. Pure Appl. Math., 15 (1878) 230–241.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, NM 55057.

SYMMETRY FACTORS FOR DIFFERENTIAL EQUATIONS

LANCE L. LITTLEJOHN
Department of Mathematics, Computer Science and Systems Design, The University of Texas at San Antonio, TX 78285

1. Introduction. It is well known (i.e. [**2**, p. 66]) that a second order linear differential equation

$$a_2(x)y''(x) + a_1(x)y'(x) + a_0(x)y(x) = 0 \tag{1}$$

can be made formally self adjoint when it is multiplied by the function

$$f(x) = \frac{e^{\int^x (a_1(t)/a_2(t))\,dt}}{a_2(x)}. \tag{2}$$

Indeed, one probably remembers that $f(x)$ is found by solving the first order equation

$$a_2(x)f'(x) + (a_2'(x) - a_1(x))f(x) = 0. \tag{3}$$

Likewise, it is a familiar fact ([**2**, p. 45], [**4**]) that an odd order differential equation with real coefficients cannot be formally self adjoint. However, the answer to the following is less well known: can a differential equation of order $2n$ ($n \geqslant 2$), with real coefficients, be made formally self adjoint by multiplying it by a suitable function $f(x)$? If so, how does one find this $f(x)$? It is the purpose of this article to answer these questions.

2. The Main Results. From here on, let $L_{2n}(y) = \sum_{k=0}^{2n} a_k(x)y^{(k)}(x)$ where a_k is real-valued, $a_k \in C^k(I)$, $a_{2n}(x) \neq 0$ for $x \in I$, where I is some compact interval of the real line R and n is some natural number. The *Lagrange adjoint* of $L_{2n}(y)$, denoted by $L_{2n}^+(y)$, is defined by

$$L_{2n}^+(y) = \sum_{k=0}^{2n} (-1)^k (a_k(x)y(x))^{(k)}.$$

$L_{2n}(y)$ is said to be *formally self adjoint* or *symmetric* when $L_{2n}(y) = L_{2n}^+(y)$. The function $f(x)$ is a *symmetry factor* for the differential expression $L_{2n}(y)$ if $f(x)L_{2n}(y)$ is formally self adjoint. (According to this definition, (2) is a symmetry factor for (1). In this case, (2) is usually called an integrating factor, but since an integrating factor is usually reserved for a function which makes a differential expression exact, we coin the term "symmetry factor.")

There are many expressions available for the most general formally self adjoint differential expression of order $2n$. For example, the reader is referred to [**1**, p. 204] and [**3**, p. 125] for two such representations. However, the one that we use here is a remarkable formula due to H. L. Krall [**4**].

THEOREM 1 (Krall). *The most general formally self adjoint differential expression of order* $2n$ *with real coefficients is given by*:

$$\sum_{k=0}^{n} b_k(x)y^{(2k)}(x) + \sum_{s=1}^{n}\sum_{k=0}^{s-1}\binom{2s}{2k+1}\frac{2^{2s-2k}-1}{s-k}B_{2s-2k}b_s^{(2s-2k-1)}(x)y^{(2k+1)}(x)$$

where B_{2i} *is the Bernoulli number defined by*:

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{i=1}^{\infty}\frac{B_{2i}x^{2i}}{(2i)!}.$$

If we replace b_k by b_{2k}, switch the order of summation in the second term and then replace k by $k-1$ in the second term, we see that the most general formally self adjoint differential expression of order $2n$ can be written as:

(4)

$$\sum_{k=0}^{n} b_{2k}(x)y^{(2k)}(x)+\sum_{k=1}^{n}\sum_{s=k}^{n}\binom{2s}{2k-1}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}b_{2s}^{(2s-2k+1)}(x)y^{(2k-1)}(x).$$

We shall prove the following theorem:

THEOREM 2. *$f(x)L_{2n}(y)$ is formally self adjoint if and only if $f(x)$ simultaneously satisfies the n homogeneous differential equations*:

$$\sum_{s=k}^{n}\sum_{j=0}^{2s-2k+1}\binom{2s}{2k-1}\binom{2s-2k+1}{j}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}a_{2s}^{(2s-2k+1-j)}(x)f^{(j)}(x) = a_{2k-1}(x)f(x),\quad k=1,2,\dots,n. \tag{5}$$

Observe when $k=n$, (5) reduces to $na_{2n}(x)f'(x)+(na'_{2n}(x)-a_{2n-1}(x))f(x)=0$. In particular, when $n=1$, we get (3). From this equation, we can determine $f(x)$ up to a nonzero constant multiple. In view of this, we restate Theorem 2.

THEOREM 3. *$f(x)$ is a symmetry factor for $L_{2n}(y)$ if and only if*:

(i)
$$f(x)=\frac{e^{1/n\int^{x}(a_{2n-1}(t)/a_{2n}(t))dt}}{a_{2n}(x)}$$

and

(ii) *for $n\geqslant 2$, $f(x)$ simultaneously satisfies the $(n-1)$ homogeneous differential equations*

$$\sum_{s=k}^{n}\sum_{j=0}^{2s-2k+1}\binom{2s}{2k-1}\binom{2s-2k+1}{j}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}a_{2s}^{(2s-2k+1-j)}(x)f^{(j)}(x)$$
$$=a_{2k-1}(x)f(x),\ \textit{for}\ k=1,2,\dots,(n-1).$$

Proof of Theorem 2. Suppose $f(x)L_{2n}(y)$ is formally self adjoint. Using (4), the functions $b_{2k}(x)=f(x)a_{2k}(x)$, $k=0,1,\dots,n$ satisfy

$$f(x)a_{2k-1}(x)=\sum_{s=k}^{n}\binom{2s}{2k-1}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}b_{2s}^{(2s-2k+1)}(x),\quad k=1,2,\dots,n$$

$$\text{i.e., } a_{2k-1}(x)f(x)=\sum_{s=k}^{n}\binom{2s}{2k-1}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}(f(x)a_{2s}(x))^{(2s-2k+1)}$$
$$=\sum_{s=k}^{n}\sum_{j=0}^{2s-2k+1}\binom{2s}{2k-1}\binom{2s-2k+1}{j}\frac{2^{2s-2k+2}-1}{s-k+1}B_{2s-2k+2}a_{2s}^{(2s-2k+1-j)}(x)f^{(j)}(x)$$

where we have used Leibniz's formula. Thus condition (5) holds. To prove the converse, retrace these steps.

In the case of constant coefficients, (5) simplifies:

$$\sum_{j=0}^{n-k}\binom{2k+2j}{2k-1}\frac{2^{2j+2}-1}{j+1}B_{2j+2}a_{2k+2j}f^{(2j+1)}(x)=a_{2k-1}f(x),\quad k=1,2,\dots,n.$$

From Theorem 3, it is necessary that the symmetry factor be $f(x)=e^{(a_{2n-1}x)/na_{2n}}$. Since

$$\frac{d^kf(x)}{dx^k}=\frac{a_{2n-1}^kf(x)}{n^ka_{2n}^k}$$

the following theorem is immediate.

THEOREM 4. *Let* $L_{2n}(y) = \sum_{k=0}^{2n} a_k y^{(k)}(x)$ *where* $a_k \in R$, $k = 0, 1, \ldots, 2n$ *and* $a_{2n} \neq 0$. *Then* $f(x) L_{2n}(y)$ *is formally self adjoint if and only if*:

(i) $$f(x) = e^{a_{2n-1}x/na_{2n}}$$

and

(ii) *for* $n \geq 2$,

$$a_{2k-1} = \sum_{j=0}^{n-k} \binom{2k+2j}{2k-1} \frac{(2^{2j+2}-1) B_{2j+2} a_{2k+2j} a_{2n-1}^{2j+1}}{(j+1) n^{2j+1} a_{2n}^{2j+1}}, \quad k = 1, 2, \ldots, (n-1).$$

3. Examples.

(1) $$L(y) = y^{(4)} - 5y^{(3)} + 5y'' + 5y' - 6y.$$

By Theorem 4, the symmetry factor is necessarily $f(x) = e^{(-5x)/2}$. However, condition (ii) of Theorem 4 is not satisfied so $L(y)$ cannot be made formally self adjoint.

(2) $$L(y) = y^{(4)} - 2y^{(3)} + y' + xy.$$

According to Theorem 3, the symmetry factor is necessarily $f(x) = e^{-x}$. It is easy to check to see that e^{-x} also satisfies condition (ii) of Theorem 3 so $e^{-x}L(y)$ is formally self adjoint. Note that

$$e^{-x}L(y) = (e^{-x}y'')'' - (e^{-x}y')' + xe^{-x}y.$$

(3) $$L(y) = x^2y^{(4)} - (2x^2 - 4x)y^{(3)} + (x^2 - (2R+6)x)y'' + ((2R+2)x - 2R)y'.$$

Using Theorem 3, it follows that e^{-x} is a symmetry factor. Again, observe that

$$e^{-x}L(y) = (x^2e^{-x}y'')'' - (([2R+2]x + 2)e^{-x}y')'.$$

(4) $$L(y) = y^{(6)} - 5y^{(5)} + 2y^{(3)} - y'' + 6y' + 3y.$$

The symmetry factor must necessarily be $f(x) = e^{-5x/3}$. However, condition (ii) of Theorem 4 is not satisfied so $L(y)$ cannot be made formally self adjoint.

Acknowledgements. The author would like to thank Professors H. L. Krall and A. M. Krall for their helpful comments and suggestions.

References

1. Coddington and Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.
2. R. H. Cole, Theory of Ordinary Differential Equations, Appleton-Century-Crofts, New York, 1968.
3. E. L. Ince, Ordinary Differential Equations, Dover, New York, 1951.
4. H. L. Krall, Self-Adjoint Differential Expressions, this MONTHLY, vol. 67, 9 (1960) 876–878.

EULER'S INTEGRALS

H. HARUKI
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

S. HARUKI
Department of Applied Mathematics, Okayama University of Science, 1-1 Ridai-cho, Okayama, 700, Japan

The aim of this note is to show an efficient method of evaluating the definite integral

(1) $$f(a) = \int_0^1 \frac{\log(x^2 - 2x\cos a + 1)}{x} dx \quad (0 \leq a \leq 2\pi).$$

CENTER SECTION

(Vol. 90, No. 7, August-September 1983)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S. All About Chess and Computers. David Levy, Monroe Newborn. Computer Sci Pr, 1982, 146 pp, $19.95. [ISBN: 0-914894-75-7] The complete works: Chess and Computers, David Levy (TR, March 1978) and More Chess and Computers, David Levy and Monroe Newborn (TR, November 1980), plus a chapter covering the four major computer chess tournaments between 1978 and 1980. The hero of these events was Ken Thompson and Joe Condon's (Bell Labs) BELLE program that performs very near Master level. LCL

General, P. Recent Trends in Mathematics, Reinhardsbrunn 1982. Ed: Herbert Kurke, et al. Teubner-Texte, B. 50. B.G. Teubner, 1982, 336 pp, (P). Papers given at or written for a conference organized by Teubner and held at Reinhardsbrunn, GDR, in October 1982. JD-B

Precalculus, T(13: 1). Technical Mathematics. Bernard J. Rice, Jerry D. Strange. Prindle, Weber & Schmidt, 1983, xi + 658 pp. [ISBN: 0-87150-327-1] Standard precalculus topics. Many examples and exercises stated in context of business and industrial situations. JRG

Precalculus, T(13: 1). College Algebra. Bodh R. Gulati. Allyn & Bacon, 1982, xviii + 567 pp, $22.95. [ISBN: 0-205-07683-1] A problem-solving approach featuring lots of exercises, examples, and applied problems. Includes chapters on exponential and logarithmic functions, matrices and determinants, sequences and series, counting. LCL

Precalculus, T(13). Algebra and Trigonometry: A Pre-Calculus Approach, Second Edition. Max A. Sobel, Norbert Lerner. Prentice-Hall, 1983, xviii + 524 pp. [ISBN: 0-13-021634-8] New edition expands emphasis on direct preparation for calculus topics. Additional exercises, graphing and word problems. Margin exercises enhance student understanding. Good readable text. (First Edition, TR, October 1979.) MW

Education, S(17-18), P. Studies in Mathematics Education, Volume 2. Ed: Robert Morris. UNESCO, 1980, x + 179 pp (P). [ISBN: 92-3-101905-8] The second in a four-volume series contains papers commissioned for a UNESCO meeting on goals in mathematics education. Investigates the impact of societal needs (especially of commerce, industry and rural development) and learner needs on the formulation of goals. Describes several national studies to examine strategies for developing, implementing and assessing goals. Conclusions and recommendations of participants are reported. MW

Education, S(16-17). SCI-MATH: Applications in Proportional Problem Solving. Madeline P. Goodstein. Addison-Wesley, 1983. Module One, 100 pp, $6 (P) [ISBN: 0-201-20072-4]; Teacher's Guide, xvi + 102 pp, $6 (P) [ISBN: 0-201-20073-2]; Module Two, 145 pp, $8.72 (P) [ISBN: 0-201-20074-0]; Teacher's Guide, xvi + 94 pp, $6 (P). [ISBN: 0-201-20075-9] Courses in problem solving by rate and ratio designed for use in science or mathematics classes. Module One is at pre-algebra level; Module Two uses algebra. Even teachers who elect not to use entire module will find good ideas in the many activities as well as applications in science and economics. MW

Foundations, P. Lecture Notes in Mathematics-940: Proper Forcing. Saharon Shelah. Springer-Verlag, 1982, xxix + 496 pp, $25 (P). [ISBN: 0-387-11593-5] Collection of lecture notes and articles on forcing methods, emphasizing their application to proving consistency and independence results. Includes results obtained prior to July 1981. KS

Foundations, S, P. Logic, Methodology and Philosophy of Science VI. Ed: L. Jonathan Cohen, et al. Stud. in Logic & Found. of Math., V. 104. Elsevier North-Holland, 1982, xiv + 856 pp, $144. [ISBN: 0-444-85423-1] Proceedings of Sixth International Congress of Logic, Methodology and Philosophy of Science, held at Hannover, West Germany, August 22-29, 1979. Theme was role of mathematics in the sciences. Divided into 14 sections: 5 on logic, 2 on history and philosophy of science, remainder on specific philosophical issues in probability, physical sciences, biology, psychology, social sciences, linguistics, scientific ethics. KS

Foundations, T(15-17: 1), S, L. Computability & Unsolvability. Martin Davis. Dover Pub, 1982, xxv + 248 pp, $6.50 (P). [ISBN: 0-486-61471-9] Reprint of 1958 edition with Davis' 1973 Monthly article on Hilbert's tenth problem as an appendix. Introduction to recursion theory using Turing machines. Covers basic theory of recursive sets and functions, logical and combinatorial decision problems, Kleene hierarchy, computable functionals. No exercises. No mathematical prerequisites except maturity. KS

Foundations, P, L. Rigorous Proof in Mathematics Education. Gila Hanna. OISE Pr, 1983, viii + 97 pp, $16.50 (P). [ISBN: 0-7744-0251-2] Full scale theoretical assault on two basic assumptions underlying the curricular reforms of the 1950's and 60's which laid new emphasis on formal rigor in the secondary curriculum. By examining the conflicting tenets of logicism, formalism and intuitionism and by looking at the social and psychological realities of actual mathematical practice, the author hopes to refute the beliefs that (a) "rigorous proof is the most important characteristic of modern mathematical practice," and (b) "in modern mathematical theory there are generally accepted criteria for the validity of a proof." GHM

Graph Theory, P. Graph Theory. Ed: Bêla Bollobás. Math. Stud., No. 62. Elsevier North-Holland, 1982, viii + 201 pp, $46.50 (P). [ISBN: 0-444-86449-0] Proceedings of conference held at Cambridge, England, March 11-13, 1981. Contains 19 research articles dealing with many aspects of graph theory. KS

Discrete Mathematics, T(14-15: 1, 2), S, L. Concepts in Discrete Mathematics. Sartaj Sahni. Camelot Pub, 1981, x + 437 pp, $26.95. [ISBN: 0-942450-00-0] Readable introduction to topics in discrete mathematics most relevant to computer science: logic, sets, relations, functions, complexity, recurrence relations, probability, graphs. No mention of Turing machines. Numerous examples and exercises; many deal with design or analysis of algorithms. KS

Number Theory, T(17: 1), S, P, L*. Sequences. H. Halberstam, K.F. Roth. Springer-Verlag, 1983, xviii + 290 pp, $28. [ISBN: 0-387-90801-3] This edition is essentially a reprint, with corrections, of the 1966 classic which gave the first systematic account of that part of number theory which deals with specific sequences of integers. (First Edition, TR, August-September 1967.) CEC

Number Theory, P. Séminaire de Théorie des Nombres, Paris 1980-81: Séminaire Delange-Pisot-Poitou. Ed: Marie -José Bertin. Progress in Math., V. 22. Birkhauser Boston, 1982, vii + 362 pp, $22.50. [ISBN: 3-7643-3066-X]

Algebra, T(15-16: 1), S, L. Introduction to Field Theory, Second Edition. Iain T. Adamson. Cambridge U Pr, 1982, viii + 181 pp, $9.95 (P); $19.95. [ISBN: 0-521-28658-1; 0-521-24388-2] Good exposition of Galois theory following Artin's approach. Includes background material on field extensions and applications to finite fields, ruler-and-compass constructions, and solution by radicals. Few but generally interesting exercises. Minor changes from First Edition. KS

Algebra, P. Lecture Notes in Mathematics-966 & 967: Algebraic K-Theory. Ed: R. Keith Dennis. Springer-Verlag, 1982. Part I, viii + 407 pp, $20 (P) [ISBN: 0-387-11965-5]; Part II, vii + 409 pp, $20 (P). [ISBN: 0-387-11966-3] Proceedings of a conference held at Oberwolfach in June 1980. JAS

Finite Mathematics, T(13). College Mathematics for the Managerial and Social Sciences. Soo Tang Tan. Prindle, Weber & Schmidt, 1983, viii + 788 pp. [ISBN: 0-87150-354-9] Standard topics in finite mathematics and calculus normally covered in a one-year course. Two years of high school algebra is the assumed background. AWR

Finite Mathematics, T(13). Finite Mathematics with BASIC: A Liberal Arts Approach. Irving Allen Dodes. Krieger Pub, 1981, xii + 359 pp, $21. [ISBN: 0-88275-862-4] A text that emphasizes linear programming, Markov chains, game theory, computation, and applications to industry--at the cost of de-emphasizing the usual work with probability, statistics, truth tables, and switching circuits. AWR

Calculus, T(13). Techniques of Calculus. Robert E. Dressler, Karl Stromberg. Amsco Coll Pub, 1983, x + 627 pp, $14 (P); $20. [ISBN: 0-87720-978-2] The standard fare for a one-year single variable calculus course intended "to teach the student how to produce correct solutions...not how to give logically pristine proofs." Is perhaps aimed at the high school calculus course. AWR

Calculus, S(13-14). Microcomputer Applications for Calculus. Gary G. Bitter. Prindle, Weber & Schmidt, 1983, xi + 241 pp, (P). [ISBN: 0-87150-378-6] A collection of very simple programs, each presented in Basic, Pascal and Fortran, illustrating most of the standard calculus topics. They do not illustrate good programming style, however. GHM

Calculus, T*(13: 1, 2). Applied Mathematics for Business and Economics, Life Sciences, and Social Sciences. Raymond A. Barnett, Charles J. Burke, Michael R. Ziegler. Dellen Pub, 1983, xvi + 856 pp, $26.95. [ISBN: 0-89517-049-3] Algebra review, finite mathematics and calculus in one and several variables. Survey-selected topics, applications and emphasis for students with one and a half to two years of high school algebra. Stress is on computational skills, ideas and problem-solving rather than on theory. No specialized experience is required for the applications. Instructor aids, including quizzes and chapter tests are available from the publisher. JK

Calculus, T(13-14: 2). Introductory Mathematical Analysis for Students of Business and Economics, Fourth Edition. Ernest F. Haeussler, Jr., Richard S. Paul. Reston Pub, 1983, xv + 904 pp. [ISBN: 0-8359-3274-5] Well-done, not overpowering treatment of calculus of one and several variables, matrix algebra, linear programming, and, in this latest edition, a chapter on basic probability theory. Self-contained with respect to applications in business and economics. Attractive format. Lots of exercises. (First Edition, TR, May 1974; Second Edition, TR, December 1976; Third Edition, TR, December 1980.) JK

Complex Analysis, S(18), P. Characteristic Properties of Quasidisks. Frederick W. Gehring. Pr U Montreal, 1982, 97 pp, $11 (P). [ISBN: 2-7606-0601-5] Quasidisks are images of the unit disk under quasiconformal mappings of the Riemann sphere. Based on lectures at the August 1982 Montreal NATO Advanced Study Institute on function theory, the book presents various characterizations of quasidisks by geometric, function-theoretic, and other properties. Some proofs and an application are given. PZ

Complex Analysis, S(18), P. Vector Valued Nevanlinna Theory. H.J.W. Ziegler. Research Notes in Math., No. 73. Pitman Pub, 1982, xiii + 201 pp, $19.95 (P). [ISBN: 0-273-08530-1] Extends the value distribution theory of F. and R. Nevanlinna to the context of simultaneous solutions of n meromorphic functions of one variable. The two main theorems of Nevanlinna theory are generalized and studied from the viewpoint of Hermitian differential geometry. With an appendix on the rudiments of complex manifolds. PZ

Complex Analysis, T(16-17: 1). Funktionentheorie. Rudolf J. Taschner. Manzsche Verlag, 1983, 241 pp, DM41.50 (P). [ISBN: 3-214-00015-2] An introduction to complex analysis written for engineers and other users of mathematics. JD-B

Differential Equations, P. Lecture Notes in Mathematics-954: Differential Systems Involving Impulses. Sudhakar G. Pandit, Sadashiv G. Deo. Springer-Verlag, 1982, vii + 102 pp, $8 (P). [ISBN: 0-387-11606-0] The derivative involved in these equations is the distributional derivative, allowing treatment of perturbed ordinary differential equations in which the perturbations are of impulsive type. This monograph attempts to unify research results of the past fifteen years. Five chapters deal with problems of existence, uniqueness, stability, boundedness, and asymptotic equivalence. AWR

Differential Equations, P. Irregular Singularities in Several Variables. A.R.P van den Essen, A.H.M. Levelt. Memoirs No. 270. AMS, 1982, iv + 43 pp, $4 (P). [ISBN: 0-8218-2270-5] Generalizes to the setting of first order linear partial differential equations a classical theorem of first order linear ordinary differential equations, to the effect that linear systems with analytic coefficients are triangulable by taking a suitable root of the independent variable, and a linear transformation of the unknown functions. PZ

Differential Equations, P. Lecture Notes in Mathematics-964: Ordinary and Partial Differential Equations. Ed: W.N. Everitt, B.D. Sleeman. Springer-Verlag, 1982, xviii + 726 pp, $32 (P). [ISBN: 0-387-11968-X] Proceedings of a conference at Dundee, Scotland, March 29-April 2, 1982 on the occasion of its centenary: 60 papers on a wide variety of research in differential equations. LAS

Differential Equations, T(15-16: 1). Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems. Richard Haberman. Prentice-Hall, 1983, x + 533 pp, $34.95. [ISBN: 0-13-252833-9] An elementary introduction to partial differential equations emphasizing engineering and physical science applications. The topics covered include separation of variables, Fourier series, Sturm-Liouville theory, Green´s functions, Fourier and Laplace transforms, and the method of characteristics. AO

Numerical Analysis, P. The Method of Discretization in Time and Partial Differential Equations. Karel Rektorys. Math. & Its Appl., V. 4. D Reidel Pub, 1982, xviii + 451 pp, $69. [ISBN: 90-277-1342-1] Generalization of the method of lines applied to elliptic, parabolic, hyperbolic, nonhomogeneous, nonlinear, and integrodifferential problems. Examples. Includes the author´s results on existence, convergence, error estimates and regularity. RWN

Numerical Analysis, P. Sparse Matrices and their Uses. Ed: Iain S. Duff. Inst. of Math. & its Appl. Academic Pr, 1981, xii + 387 pp, $43.50. [ISBN: 0-12-223280-1] Proceedings of a conference held July 9-11, 1980 at the University of Reading (England). Most of the papers are accessible to the nonspecialist. AO

Functional Analysis, P. Topics in Locally Convex Spaces. Manuel Valdivia. Math. Stud., V. 67. Elsevier North-Holland, 1982, xiii + 510 pp, $69.75 (P). [ISBN: 0-444-86418-0] Aimed at people who are already familiar with the general theory of locally convex spaces, the book collects recent results from the periodical literature and sets forth some new results. AWR

Functional Analysis, P. Lecture Notes in Mathematics-955: Bundles of Topological Vector Spaces and Their Duality. Gerhard Gierz. Springer-Verlag, 1982, iv + 296 pp, $13.50 (P). [ISBN: 0-387-11610-9]

Functional Analysis, T(18: 1), S, P. Linear Equations in Banach Spaces. S.G. Krein. Transl: A. Iacob. Birkhauser Boston, 1982, xii + 102 pp, $14.95. [ISBN: 3-7643-3101-1] A brief but comprehensive treatment of linear equations in Banach spaces. Covers, e.g., solvability, stability, special

classes of equations, the adjoint equation, regularity. Concludes with applications to integral and differential operators. Appendix on elementary functional analysis; no exercises. Russian edition appeared in 1971. PZ

Functional Analysis, P. Spectral Theory. Ed: Wieslaw Zelazko. Banach Center Pub, V. 8. PWN, 1982, 603 pp. [ISBN: 83-01-01495-4] Survey papers, summaries of lectures, and research contributions from the 1977 spectral theory semester at the Banach Center in Warsaw. Concludes with a collection of problems from the Center's problem book. LAS

Functional Analysis, T(18: 1), S, P. The Convolution Product and Some Applications. Wilhelm Kecs. Transl: Victor Giurgiutiu. D Reidel Pub, 1982, xvii + 332 pp, $69.50. [ISBN: 90-277-1409-6] A treatment of the basic properties and applications of the convolution product for functions and distributions. After a general introduction to topological vector spaces and convolutions, applications are made to integral transforms, differential equations, electrical engineering, and viscoelastic solids. References, index. JS

Functional Analysis, P. Der Graphensatz in lokalkonvexen topologischen Vektorräumen. Michael Oberguggenberger. Teubner-Texte zur Math., B. 44. B.G. Teubner, 1982, 148 pp, 15,50M (P). On Banach's closed graph theorem and some of its generalizations. JD-B

Analysis, P. Nonlinear Analysis, Function Spaces and Applications, Volume 2. Ed: Oldrich John, Alois Kufner. Teubner-Texte, B. 49. B.G. Teubner, 1982, 268 pp, 27M (P). Proceedings of the Spring School held in Písek, Czechoslovakia, May 24-28, 1982. JAS

Algebraic Geometry, S(18), P. Commutative Algebra: Proceedings of the Trento Conference. Ed: Silvio Greco, Giuseppe Valla. Lect. Notes in Pure & Appl. Math., V. 84. Dekker, 1983, viii + 351 pp, $47.50 (P). [ISBN: 0-8347-1899-2] Collection of research papers representing an overview of the most recent developments in commutative algebra with special reference to algebraic geometry. LCL

Differential Geometry, T(18: 1), P. Total Curvature in Riemannian Geometry. T.J. Willmore. Ser. in Math. & Its Applic. Halsted Pr, 1982, 168 pp, $39.95. [ISBN: 0-470-27354-2] An exposition of "the invariants which arise from integrating various curvature measures over manifolds." Very definitely a sophisticated presentation of the material. No exercises, good bibliography, a short and rather incomplete index. JAS

Differential Geometry, P. Differential Geometry and Mathematical Physics. Ed: M. Cahen, et al. Math. Physics Stud., No. 3. D Reidel Pub, 1983, vii + 188 pp, $28.50 (P). [ISBN: 90-277-1508-4] Lectures given at the meetings of the Belgian Contact Group on Differential Geometry held at Liège, May 2-3, 1980 and at Leuven, February 6-8, 1981. JAS

Geometry, T(16-17), P. Combinatorial Integral Geometry With Applications to Mathematical Stereology. R.V. Ambartzumian. Wiley, 1982, xvii + 221 pp, $45. [ISBN: 0-471-27977-3] Could as easily be described as a book on probability or combinatorics. A fresh development built upon the idea of introducing measures into the space of lines in the plane to solve problems in geometrical probability. AWR

Geometry, T(17:1, 2), S, P. An Introduction to Convex Polytopes. Arne Brønsted. Grad. Texts in Math., No. 90. Springer-Verlag, 1983, viii + 160 pp, $28. [ISBN: 0-387-90722-X] Assumes only elementary linear algebra and some point set topology; develops basic material on d-dimensional convex sets and, particularly, convex polytopes with the goal of studying the combinatorial theory of convex polytopes. Highlights are the Dehn-Sommerville relations, upper bound theorem and lower bound theorem, the latter two dating from the early 1970's. SS

Geometry, P. Lecture Notes in Mathematics-970: Twistor Geometry and Non-Linear Systems. Ed: H.D. Doebner, T.D. Palev. Springer-Verlag, 1982, v + 216 pp, $11.50 (P). [ISBN: 0-387-11972-8] Survey lectures from the Fourth Bulgarian Summer School on Mathematical Problems of Quantum Field Theory held at Primorsko, Bulgaria in September 1980. JAS

Operations Research, T(14: 1). Introduction to Mathematical Programming: Quantitative Tools for Decision Making. Benjamin Lev, Howard J. Weiss. Elsevier North-Holland, 1982, xii + 289 pp, $27.50. [ISBN: 0-444-00591-9] Intended for graduate or undergraduate business students. Few mathematics prerequisites, but does require some sophistication. Topics include linear programming, application to the classical transportation and assignment problems, integer and dynamic programming, search techniques in non-linear programming. JRG

Operations Research, P. Mathematical Programming with Data Perturbations II. Ed: Anthony V. Fiacco. Lect. Notes in Pure & Appl. Math., V. 85. Dekker, 1983, vi + 155 pp, $34.50 (P). [ISBN: 0-8247-1789-9] Papers selected from a May 1980 symposium at George Washington University covering recent results in programming sensitivity and stability analysis methodology. LAS

Game Theory. The Mathematics of Games of Strategy: Theory and Applications. Melvin Dresher. Dover Pub, 1981, 184 pp, $4 (P). [ISBN: 0-486-64216-X] An unabridged republication of the author's book Games of Strategy: Theory and Applications, Prentice-Hall, 1961. Written to an audience that has had a year of calculus but no linear algebra, this is a self-contained exposition that starts with parlor games, proves the minimax theorem, and discusses a variety of ways to find optimal strategies. A bit more than a third of the book treats infinite games. Explanation is largely couched

in military jargon. Attractive as a text except for lack of exercises. AWR

Optimization, S(16-17), P, L. Nonlinear Mathematics. Thomas L. Saaty, Joseph Bram. Dover Pub, 1981, xv + 381 pp, $6.50 (P). [ISBN: 0-486-64233-X] Reproduction of this same title published in 1964 by McGraw-Hill. The goal is to provide some unified approaches to nonlinear mathematics. Attention is given to nonlinear problems in optimization, programming, ordinary differential equations, automatic control, and prediction theory. AWR

Optimization, T(17-18: 2), P. Optimization--Theory and Applications: Problems with Ordinary Differential Equations. Lamberto Cesari. Appl. in Math., No. 17. Springer-Verlag, 1983, xiv + 542 pp, $68. [ISBN: 0-387-90676-2] Focused on nonparametric problems of the calculus of variations in one independent variable and on problems of optimal control modelled by ordinary differential equations, this book emphasizes the connections of present work to classical theory. Numerous concrete problems appear in the text before existence theorems are taken up. Intended for graduate students in mathematics or engineering. AWR

Optimization, T(16-17). Practical Optimization. Philip E. Gill, Walter Murray, Margaret H. Wright. Academic Pr, 1981, xvi + 401 pp, $46.50 (P). [ISBN: 0-12-283952-8] Emphasis is on implementation of optimization methods with particular attention to the effects of finite-precision computation. To make the book self-contained and more suitable as a text, chapters are included on numerical linear algebra and on a treatment of optimality conditions. Authors have all been active in development of software for the solution of optimization problems. AWR

Optimization, S(17-18), P. Grundprinzipien der Theorie der Extremalaufgaben. Vladimir M. Tichomirov. Teubner-Texte zur Math., B. 30. B.G. Teubner, 1982, 152 pp, 16M (P). A brief and modern introduction to the general theory of extremal problems. JD-B

Optimization, P. Constrained Optimization and Lagrange Multiplier Methods. Dimitri P. Bertsekas. Computer Sci. & Appl. Math. Academic Pr, 1982, xiii + 395 pp. [ISBN: 0-12-093480-9] This monograph presents the method of multipliers and many of its variations that have been developed in the past fifteen years. Several previously unpublished results and algorithms are given. AO

Optimization, T(16-18: 1). Matrices and Simplex Algorithms. A.R.G. Heesterman. D Reidel Pub, 1983, x + 790 pp, $96. [ISBN: 90-277-1514-9] After an elementary introduction to matrices and determinants, the emphasis is on programming algorithms. Aimed largely toward applications in economics, it deals with linear, quadratic, and integer programming. Numerous examples, exercises, and programs (in Algol). References, index. Inordinately priced. JS

Statistics, P. Lecture Notes in Statistics-12: Statistical Analysis of Counting Processes. Martin Jacobsen. Springer-Verlag, 1982, vii + 226 pp, $14.80 (P). [ISBN: 0-387-90769-6] A counting process is a stochastic process that counts the number of events that have occurred before time t. This monograph treats multiplicative intensity models for counting processes first introduced by Odd Aalen in 1975. LAS

Statistics, S(16-18), P. Lecture Notes in Medical Informatics-19: Discovery and Representation of Causal Relationships from a Large Time-Oriented Clinical Database: The RX Project. Robert L. Blum. Springer-Verlag, 1982, xix + 242 pp, $18 (P). [ISBN: 0-387-11962-0] RX is a computer program that derives from a clinical database a set of causal conjectures. The principles behind RX can be used in any area of knowledge to draw inferences (rather than merely retrieve information) from a database. LAS

Statistics, S(15-18), P. Lecture Notes in Statistics-15: Sampling With Unequal Probabilities. K.R.W. Brewer, Muhammad Hanif. Springer-Verlag, 1983, ix + 164 pp, $12.80 (P). [ISBN: 0-387-90807-2]

Statistics, S(16-18), P. Regression Analysis of Survival Data in Cancer Chemotherapy. Walter H. Carter, Jr., Galen L. Wampler, Donald M. Stablein. Statistics, V. 44. Dekker, 1983, x + 209 pp, $35. [ISBN: 0-8247-1736-8] An exhortation, supported by extensive preclinical evidence, to the use of regression models in clinical trials of cancer drugs, especially in predicting optimal doses when several drugs are used at once. LAS

Statistics, P. COMPSTAT 1982, Part 1: Proceedings in Computational Statistics. Ed: H. Caussinus, P. Ettinger, R. Tomassone. Physica-Verlag, 1982, 466 pp, (P). [ISBN: 3-7051-0002-5] 15 invited and 56 contributed papers from a 1982 Toulouse symposium on computational statistics providing a diverse, international perspective on this rapidly growing field. LAS

Statistics, S(15-16). Initiation à l'Analyse des Données. Jean de Lagarde. Dunod (UD Distr: SMPF, 485 5th Ave., Suite 1042, NY 10017), 1983, ix + 157 pp, 68F (P). [ISBN: 2-04-105493-0] An introduction to statistics, intended for users and without proofs. Several programs in Basic. JD-B

Computer Literacy, S. What's So Funny About Computers? S. Harris, William Kaufmann, 1982, vi + 122 pp, $6.95 (P). [ISBN: 0-86576-049-7] 120 antidotes to "terminal shock," all with some byte to them. "The central processing unit is on strike. How does one negotiate with a central processing unit?" LAS

Computer Programming, S. Fast BASIC: Beyond TRS-80 BASIC. George A. and Thomas G. Gratzer. Wiley, 1982, ix + 278 pp, $14.95 (P). [ISBN: 0-471-09849-3] Describes how to increase speed of Basic programs by using peek and poke commands and machine language routines from TRS-80 ROM. Contains background information on representation of numbers, memory organization and Z-80 microprocessor. Exercises with complete solutions. KS

Computer Programming, T(13-14: 1), S, L. Programming Concepts and Problem Solving: An Introduction to Computer Science Using Pascal. Peter Linz. Ser. in Comput. & Inform. Sci. Benjamin/Cummings, 1983, xxii + 388 pp, $21.95. [ISBN: 0-8053-5710-6] Introduction to programming using Pascal. Suitable for a first course, possibly a year course if supplemented. RM

Computer Programming, T(13: 1). Introduction to Programming BASIC: A Structured Approach, Second Edition. Chris R. Siragusa. Prindle, Weber & Schmidt, 1983, xiii + 317 pp (P). [ISBN: 0-87150-386-7] Structured approach to programming introduced early and used throughout the text. Examples and exercises require minimal degree of mathematical sophistication. Changes from First Edition include changes in the order in which Basic instructions are introduced, new techniques for data retrieval, more emphasis on programming style. JRG

Computer Programming, T(13: 1). Introduction to BASIC Programming. Gary B. Shelly, Thomas J. Cashman. Anaheim Pub, 1982, xiii + 404 pp, (P) [ISBN: 0-88236-118-X]; Test Bank, 153 pp; Transparency Masters, 219 pp. Detailed introduction, with stress on the importance of good program design, proper documentation, and readable code. Abundant advice, coding tips, examples, review questions, exercises, summaries. Unusually fine color graphics on high enamel stock. The Test Bank contains nearly 100 true-false questions and 500 multiple-choice questions. All the drawings in the text are included in Transparency Masters. LCL

Computer Programming, S*(13). Introduction to Structured Programming with Pascal. Milton M. Underkoffler. Prindle, Weber & Schmidt, 1983, ix + 279 pp, (P). [ISBN: 0-87150-394-8] An introduction to Pascal which emphasizes top-down design, structured programming techniques, and program documentation. Thirty-three completely solved problems illustrate the main points made in the text. AO

Computer Programming, S(15-17), P, L. Studies in Ada Style, Second Edition (January 1983). Peter Hibbard, et al. Springer-Verlag, 1983, 111 pp, $12 (P). [ISBN: 0-387-90816-1] The first part of this monograph presents some of the ideas that motivated the development of Ada. The second part presents five Ada programs together with discussion of their design. AO

Computer Programming, S(15-17), P. ADA, An Introduction, Second Edition--January 1983. Henry Ledgard. Springer-Verlag, 1983, viii+ 135 pp, $12.80 (P). [ISBN: 0-387-90814-5] The text has been slightly revised to bring it into conformity with the July 1982 Ada standard. (First Edition, TR, February 1982.) AO

Computer Programming, S(13). Programming the IBM Personal Computer: FORTRAN 77. Robert A. Rouse, Thomas L. Bugnitz. Holt, Rinehart & Winston, 1983, xiv + 240 pp, $16.95 (P). [ISBN: 0-03-062042-2] An elementary introduction to Fortran 77 emphasizing structured programming techniques. The text is geared specifically to the Fortran 77 available on the IBM PC. AO

Computer Programming, T(13: 1), S. FORTRAN Programming: A Spiral Approach, Second Edition. Charles B. Kreitzberg, Ben Shneiderman. HarBrace J, 1982, x + 437 pp, $17.95 (P). [ISBN: 0-15-528015-5] An elementary introduction to Fortran (including some of the features of Fortran 77) compatible with WATFOR/WATFIV. Each chapter includes a reading on programming style and discusses debugging techniques. Structured flowcharts are used to present algorithms. AO

Data Structures, S(13-14). Elementary Computer Graphics. Aftab A. Mufti. Reston Pub, 1983, xiv + 210 pp, $19.95. [ISBN: 0-8359-1654-5] A very sketchy survey of hardware, data structures, and matrix operations associated with two and three dimensional computer graphics, illustrated primarily by wire-frame drawings and associated Fortran programs. More of an outline than a monograph, it just skims the surface of a fascinating subject. LAS

Computer Science, T(15-17: 1), L. Data Models. Dionysios C. Tsichritzis, Frederick H. Lochovsky. Prentice-Hall, 1982, xiv + 381 pp, $24.95. [ISBN: 0-13-196428-3] Data models are used in the design of database management systems to organize and represent information in a form amenable to computer manipulation. This introductory text describes a number of data models and illustrates their use. AO

Computer Science, T(15-17: 1), S, L. Abstract Machines and Grammars. Walter J. Savitch. Little, Brown & Co, 1982, xiv + 215 pp, $24.95. [ISBN: 0-316-771619] Concise introduction to computability, automata and formal language theory. Good motivation of topics and discussion of relations between topics. Assumes familiarity with high-level language like Pascal. KS

Computer Science, T(14-18: 1, 2), S, L. Fundamentals of Programming Languages. Ellis Horowitz. Computer Sci Pr, 1983, xiv + 450 pp, $23.95. [ISBN: 0-914894-37-4] Focuses on topics and concepts found in most programming languages: syntax, variables, expressions, statements, typing, scope, procedures, data types, data abstraction, exception handling, concurrency, I/O. Contains chapters on functional programming, data flow programming, and object oriented programming languages. Each chapter ends with a list of concepts discussed in the chapter and a set of chapter exercises. References, index. RJA

Computer Science, S(15-18), P, L. History of Programming Languages. Ed: Richard L. Wexelblat. Academic Pr, 1981, xxiii + 758 pp, $45. [ISBN: 0-12-745040-8] Proceedings of the ACM SIGPLAN History of Programming Languages Conference, June 1-3, 1978. Contains an editor's introduction, an afterword, and an index. RJA

Computer Science, S(15-18), P. The Architecture of High Performance Computers. Roland N. Ibbett. Springer-Verlag, 1982, 172 pp, $12.50 (P). [ISBN: 0-387-91215-0] An historical approach to the ideas of computer architecture via case studies of machines from the CDC 6600 atlas, MU5, and others in the 60's, to current supercomputers such as the CRAY-1. JAS

Computer Science, L. International Microcomputer Dictionary. Sybex, 1981, x + 121 pp, $3.95 (P). [ISBN: 0-89588-067-9] It really is! Sixty-one pages of dictionary are followed by a list of common chip numbers and nearly forty pages of Danish, Dutch, French, German, Hungarian, Italian, Norwegian, Polish, Spanish and Swedish indexed by the English term. The book concludes with specifications on such things as the IEEE-488 Bus and a list of microcomputer companies. Definitely $3.95 worth of information. JAS

Computer Science, S. A Practical Introduction to Pascal, Second Edition Including the Pascal Standard. I.R. Wilson, A.M. Addyman. Springer-Verlag, 1982, xii + 239 pp, $14 (P). [ISBN: 0-387-91210-X] A tutorial based on lectures at the University of Manchester, and benefiting from the authors' work on the Pascal Standard. Has both exercises (pencil and paper) and problems (programming). Answers to exercises and programs for selected problems. Also includes valuable 70 -page small type reproduction of the BS6192:1982 specification for Pascal. (First Edition, TR, June-July 1979.) RBK

Control Theory, P. Lecture Notes in Mathematics-963: Optimal Processes on Manifolds; an Application of Stokes' Theorem. Roel Nottrot. Springer-Verlag, 1982, vi + 124 pp, $8 (P). [ISBN: 0-387-11963-9] A process is a solution of differential equations which depend on parameters that control the process. This monograph uses Stokes' theorem to develop a maximum principle for optimal processes defined on manifolds, thereby reducing the problem of optimal control to a problem of mathematical programming. LAS

Control Theory, P. Differential Geometric Control Theory. Ed: Roger W. Brockett, Richard S. Millman, Hector J. Sussmann. Progress in Math., V. 27. Birkhauser Boston, 1983, vii + 340 pp, $25. [ISBN: 3-7643-3091-0] Differential geometric methods applied to nonlinear control theory; proceedings of Michigan Technological Conference, Summer, 1982. Two expository series, R. Gardner on exterior differential forms and Sussmann on control theory, constitute over half. RB

Systems Theory, P. Lecture Notes in Control and Information Sciences-41: Nonlinear Time-discrete Systems: A General Approach by Nonlinear Superposition. M. Gössel. Springer-Verlag, 1982, 115 pp, $8 (P). [ISBN: 0-387-11914-0] Chapter headings include: automata-definitions and notations, linear automata, automata superponable with respect to pairs of operations. AWR

Applications (Dynamics), S(14-16), L. Dynamics: The Geometry of Behavior. Part One: Periodic Behavior. Ed: Ralph H. Abraham, Christopher D. Shaw. Visual Math. Lib. Vismath V. 1. Aerial Pr, 1982, x + 220 pp, $25 (P). [ISBN: 0-942344-01-4] First of a four-part series consisting of captioned four-color illustrations of the mathematical ideas of nonlinear dynamics: links periodic behavior (simple, forced, and coupled oscillations) in physical systems to trajectory pictures on the plane and the torus. Based on the authors' computer graphics experience with the Visual Math Project; companion Apple disks are available. JNC

Applications (Economics), T(16-17), L. Lecture Notes in Economics and Mathematical Systems-205: Introduction to the Theory of Economic Growth. Ramu Ramanathan. Springer-Verlag, 1982, ix + 347 pp, $23 (P). [ISBN: 0-387-11943-4] Distinguished from other books on the same topic, according to the author, by being written with students, not professional colleagues in mind. Presupposes intermediate micro and macro economics and a working knowledge of calculus. Various models are all cast in the same analytic framework, and care is taken to spell out the underlying assumptions. Exercises follow each chapter. AWR

Applications (Economics), P. Current Developments in the Interface: Economics, Econometrics, Mathematics. Ed: M. Hazewinkel, A.H.G. Rinnooy Kan. D Reidel Pub, 1982, ix + 355 pp, $44. [ISBN: 90-277-1505-X] Proceedings of a January 1982 symposium in Rotterdam in honor of the 25th anniversary of the Econometric Institute of the Erasmus University, intended to cross-fertilize economics, mathematics, and econometrics. Includes 13 chapters, each preceded by brief introductory remarks and followed by summaries of questions and comments. LAS

Applications (Engineering), S(14-15), P, L.** Instabilities and Catastrophes in Science and Engineering. J.M.T. Thompson. Wiley, 1982, xvi + 226 pp, $34.95. [ISBN: 0-471-09973-2] A discussion of how a stable equilibrium may become unstable, how a continuous process may become discontinuous. Such phenomena are cited in engineering structures, exploding stars, the fracture of a crystal lattice, turbulence in fluid flow, and more. Written in the style of Scientific American, this book will be undergirded by a companion volume Static and Dynamic Instability Phenomena, by J.M.T. Thompson and G.W. Hunt, to be published by Wiley. AWR

Applications (Engineering), T(15). Advanced Engineering Mathematics. Peter V. O'Neil. Wadsworth Pub, 1983, xvi + 1211 pp, $36.95. [ISBN: 0-534-01136-5] Everything the engineer needs to know about

post-calculus mathematics, exclusive of computing: differential equations (including Laplace transforms, special functions), vectors and matrices, vector analysis, Fourier analysis and boundary value problems, complex analysis, and numerical methods. AWR

Applications (Engineering), S(17-18), P. Robot Vision. Ed: Alan Pugh. Springer-Verlag, 1983, xi + 356 pp, $47.50. [ISBN: 0-387-12073-4] The first in a series concerning manufacturing applications of microprocessor controls and robotics. This volume contains a collection of 28 papers ranging from two general introductory papers to technical papers on specific industrial machines and applications, e.g., automatic chocolate decoration. JAS

Applications (Engineering), T(16-17: 1, 2), L. The Finite Element Method for Engineers, Second Edition. Kenneth H. Huebner, Earl A. Thornton. Wiley, 1982, xxii + 623 pp, $37.50. [ISBN: 0-471-09159-6] A broad, somewhat self-contained introduction. Foundations from matrix theory, interpolation, calculus of variations and physics. Application to elasticity, fluid mechanics, heat transfer, lubrication and general field problems. Includes some computer codes, good problems and bibliographies. RWN

Applications (Engineering), P*. Applied Mathematical Analysis: Vibration Theory. Ed: G.F. Roach. Math. Ser., No. 4. Birkhauser Boston, 1982, 229 pp, $29.95. [ISBN: 0-906812-12-7] Seventeen papers on various aspects of vibration theory ranging from geometrical optics and acoustic scattering to the mathematical modelling of a sheet rolling mill. Proceedings of the first annual seminar on open questions in applied mathematical analysis at the University of Strathclyde. JK

Applications (Engineering), P. Fast Transforms: Algorithms, Analyses, Applications. Douglas F. Elliott, K. Ramamohan Rao. Academic Pr, 1982, xxii + 488 pp. [ISBN: 0-12-237080-5] This monograph presents many of the fast transform algorithms (e.g., discrete Fourier, Walsh-Hadamard) of importance in applied engineering practice as well as analyses of their properties and some of their applications. AO

Applications (Epidemiology), P. Environmental Epidemiology: Risk Assessment. Ed: Ross L. Prentice, Alice S. Whittemore. SIAM, 1982, ix + 229 pp, $27.50 (P). [ISBN: 0-89871-184-3] Proceedings of a 1982 summer conference at Alta, Utah, sponsored by the SIAM Institute for Mathematics and Society (SIMS). LAS

Applications (Management), T(14-16: 1, 2), S, L. Introduction to Data Base Management in Business. James Bradley. Holt, Rinehart & Winston, 1983, ix + 630 pp, $31.95. [ISBN: 0-03-061693-X] A readable and quite thorough presentation with no specific prerequisites other than general economics and elementary programming (at the most). Lots of discussion questions. The emphasis is on large commercial systems so there are few if any programming problems for minis or micros. JAS

Applications (Physics), P. Lecture Notes in Physics-169: Constrained Dynamics. Kurt Sundermeyer. Springer-Verlag, 1982, iv + 318 pp, $13 (P). [ISBN: 0-387-11947-7] A development of constrained Hamiltonian systems, also known as "singular Lagrangian systems" with applications to Yang-Mills theory, general relativity, classical spin, and the dual string model. Contains a number of appendices with useful mathematical background material--differential and Riemannian geometry, Feynman path-integrals, and the like. JAS

Applications (Social Science), T(13-14: 1, 2), L. Mathematics with Applications in the Management, Natural, and Social Sciences, Third Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1983, 690 pp, $21.95. [ISBN: 0-673-15793-8] The changes in the Third Edition of this well-known text (First Edition, TR, November 1974; Extended Review, June-July 1975; Second Edition, TR, April 1979) include: expansion of the chapters on review of algebra, functions and calculus, rewriting of the chapter on mathematics of finance; and reorganization of material on sets and counting and on advanced probability topics. RB

Applications (Social Science), S(15-16), P, L? Concepts and Models of a Quantitative Sociology: The Dynamics of Interacting Populations. W. Weidlich, G. Haag. Ser. in Synergetics, V. 14. Springer-Verlag, 1983, xii + 217 pp, $31.50. [ISBN: 0-387-11358-4] Synergetics, the study of situations in which a system composed of many parts or individuals acquires a new structure on macroscopic scales, has been developed largely in physics, chemistry, and biology. This monograph seeks to carry the message into the social sciences. The first three chapters introduce the idea, the last three seek to apply it to population dynamics, the theory of investment, and the interaction of competitive macrosocieties. AWR

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Oklahoma-Arkansas Section

The fall meeting of the Oklahoma-Arkansas Section was held at the University of Oklahoma on March 18-19, 1983. The attendance was approximately 100 people with 23 papers given by faculty and students.

Invited Address:

"Some Mathematical Surprises," by Victor Klee, University of Washington. (N.A. Court Lecture)
"Report on the National Science Board Commission and Precollege Education in Mathematics," by Katherine Layton, National Science Board Commission on Precollege Mathematics, Science, and Technology.

Panel Discussion:

"Precollege Mathematics Education in the United States," Katherine Layton (Presider); R.D. Anderson, Past President of MAA; Dale Franks, Superintendent of the Hope Public School System, Hope, Arkansas; Dave McCurdy, United States Congressman, 4th District Representative, Oklahoma; and Charles Watson, Mathematics Specialist, Department of Education, Arkansas.

Short Presentations:

"Continued Square Roots," by Bianca M. Hearn, Hendrix College.
"p-adic Numbers," by Daryl Ezzo, Oral Roberts University.
"A Child's Garden of Quaternions," by Bobby Winters, East Central University.
"Elementary Transfer Orbits," by Tim Koster, Oklahoma State University.
"Log-Convexity and the Gamma Function," by Karen Anderson, Hendrix College.
"Functional Equations Arising from Notational Ambiguities in Calculus," by Karen Shirley, Hendrix College.
"Two Ancient Greek Construction Problems in Euclidean and Hyperbolic Geometry," by Jack M. Rau, Oklahomna State University.
"The Diffusion Index as an Indicator of Stock Market Trends," by Dennis Bertholf, Harry Comeskey, and Wayne B. Powell, Oklahoma State University.
"Robust Autoregressive Time Series Parameter Estimation," by John Fingerlin, G.O. Scan-AMCO.
"Digital Image Processing," by Frank Chimenti, Southwestern State University.
"Some Closed Graph Theorems," by Ray Hamlett, East Central University.
"Problem Solving," by Doug Foster, Oral Roberts University.
"A one-hour course on Hofstadter's Godel, Escher, Bach," by Dale Alspach, Michael Folk, and Joel Haack, Oklahoma State University.
"A Mathematical System of Single Fingerprint Classification," by William R. Orton, University of Arkansas at Fayetteville.
"An Applied Linear Algebra Course," by Tom Cairns and Bill Coberly, University of Tulsa.
"Functions of Mathematics Augmented With a Micro," by Kelvin Casebeer, Southwestern State University.
"Survey of the Mathematics Shortage in Oklahoma," by Harold Huneke, University of Oklahoma.
"Problem Solving for Gifted Junior High Students," by Paul Duvall, Oklahoma State University.
"Apportionment: A Model with Means," by Donald L. Pattern, University of Oklahoma.
"Joint Graphs and the Traversability," by Mahesh M. Hiremath, University of Arkansas at Fayetteville.
"Supercyclic Hilbert Space Operators," by Bill Stockwell, Central State University.
"Constructing Amicable Pairs From Known Amicable Pairs: An Alternative Approach," by Dale Woods and Joe D. Flowers, Central State University.
"Maximum Likelihood Estimation of Linear Sufficient Statistics," by Bill Coberly, University of Tulsa.

Nebraska Section

The fall meeting of the Nebraska Section was held at the University of Nebraska, Omaha on March 25-26, 1983.

Invited Lecture:

"MAA's New Mathematical Science Curriculum Recommendations," by Alan Tucker, SUNY at Stony Brook.

Contributed Papers:

"Numerical Estimate of Shock Region of a Certain Hyperbolic 2-Conservation Law," by Peter H. Chang, University of Nebraska, Omaha.

"Modular Arithmetic, Second Differences and Symmetry," by Mel Thornton, University of Nebraska, Lincoln.

* "Objective Testing with Subjective Probability," by Nelson C. Fong, Kearney State College.

"Good Principles Gone Awry: How Not to Teach Programming," by Albert F. Shpuntoff, University of Nebraska, Omaha.

"Exceptional t-Designs," by Sypros S. Magliveras and David W. Leavitt, University of Nebraska, Lincoln.

* "High School Math Contest," by Richard Barlow, Kearney State College.

"Paradoxical Coverings of the Real Line," by Ivan Niven, President, Mathematical Association of America.

"Group Codes for the Gaussian Channel: A Survey of Current Research," by John Karlof, University of Nebraska, Omaha.

"On the Number of Combinations Without r-Separation," by John Konvalina, University of Nebraska, Omaha.

"Conditional Yeh-Wiener Integrals," by Kun Soo Chang, University of Nebraska, Lincoln.

"How to Solve the nXnXn Cube," by J.A. Eidswick, University of Nebraska, Lincoln.

"To Compute or Not to Compute: The KSW Computer Science Placement Test," by Stanley Wileman, University of Nebraska, Omaha.

"A Basic Maximum Flow Algorithm for Network Modules of Building Evacuation," by John T. Hsieh, University of Nebraska, Omaha.

Illinois Section

The sixty-second annual spring meeting of the Illinois Section convened at Rockford College on April 29-30, 1983 with approximately seventy members in attendance.

Invited Addresses:

"Problems of Intuitive Geometry," by Kenneth Stolarsky, University of Illinois.

"The Logic of Differential Equations," by Lee Rubel, University of Illinois.

"Modern Applications of Second Grade Mathematics," by Charles Vanden Eynden, Illinois State University.

"Root Systems and Arrangements of Hyperplanes," by Louis Solomon, University of Wisconsin.

"What Are the Waves of the Future?" by Carole Bauer, Triton College.

"Mathematics and Axiomatic Social Choice Theory: The Arrow (Theorem) Has Left Its Mark," by Ed Packel, Lake Forest College.

"A Mathematical Principle for the Circumvention of Multiple-Choice Examinations," by Graham Evan, University of Illinois.

"CUPM's New Recommendations for a General Mathematical Sciences Program," by Alan Tucker, SUNY at Stony Brook.

"A Random Talk," by Robert V. Hogg, University of Iowa.

Professor Howard C. Saar, Secretary-Treasurer of ISMAA for the past seventeen years, was presented the Distinguished Service Award of the Section.

Seaway Section

The Seaway Section held its spring meeting jointly with the New York State Mathematics Association of Two Year Colleges at Utica, New York on April 22-24, 1983. Over 200 mathematicians were in attendance.

Invited Addresses:

"The Mathematics Teacher of the 1980's," by Stanley Bezuszka, Boston College.

"Some Recent Applications of Functional Equations to the Behavioral Sciences" (The Gehman Lecture), by Janós D. Aczél, University of Waterloo.

Short Papers:

"Fagin's Theorem on Domain-Key Normal Form for Relations," by Douglas Cashing, St. Bonaventure University.

"On the Use of Multivariable Polynomials for Integrating Ordinary Differential Equations," by Mike Mikalajanus, Montreal, Quebec.

"Writing About Mathematics," by Dorothy Buerk, Ithaca College.

"Simple Sum-Preserving Rearrangements of Series: A Discouraging Theorem," by Paul Schaefer, SUNY College at Geneseo.

"Developments from Normal Surface Theory which Apply to Knot Spaces," by Richard Gustafson, SUNY College at Oneonta.

"Business Simulation Games: A Successful Interim Course," by Patti Frazer Lock, St. Lawrence University.

Student Papers:

"Rubik Type Puzzles and Their Groups," by Karen F. Gold, SUNY at Albany.
"Computer Graphics Generation of Wallpaper Groups," by John Gennavi, Nancy Goering, and Keith Swingruber, Colgate University.
"The Solution to Bulgarian Solitaire," by Emilio Mastrandrea, Rochester Institute of Technology.

It was announced that, for the second year, David Ash, University of Waterloo, is the Putnam Prize winner from the Seaway Section.

Rocky Mountain Section

The sixty-sixth annual spring meeting of the Rocky Mountain Section was held on April 29-30, 1983 on the campus of Colorado State University in Ft. Collins, Colorado. There were 125 MAA members in attendance.

Invited Lecture:

"Mathematics in 1983: Our Problems, Our Prospects, and our Constituency," by Gail Young, University of Wyoming.

Panel Discussion:

"Implications of MuMath on the Mathematics Curriculum," by A.R. Brown, Colorado School of Mines; William Dorn, University of Denver; Darel Hardy, Colorado State University; Gail Young, University of Wyoming.

Short Presentations:

"Legal Protection of Computer Software: A Glitch in the System," by Antonette Logar, National College.
"A Two-Year Curriculum Integrating Discrete and Continuous Mathematics," by Ronald E. Prather, University of Denver.
"Edge-labelled Trees," by Julie Yancey, Fort Lewis College.
"Splittings of a Definite Integral," by Hung Li, University of Southern Colorado.
"A Senior Design Course in Computer Science From the Students' Point of View," by Colleen Borstad, South Dakota School of Mines and Technology.
"Finite Math for the Freshman Computer Science Student," by Karen Whitehead, South Dakota School of Mines and Technology.
"Observable Differences Between Male and Female Computer Science Students," by James Sandau and Sheri Kirley, South Dakota School of Mines and Technology.
"Demonstration of MuMath," by Aaron Meyerowitz, Colorado State University.
"Seven Notable Women Mathematicians," by Julia Ann Walker, Boulder, Colorado.
"The Jordan Curvve Theorem," by David Lear, University of Colorado.
"A Mathematical Model Predicting Ultimate Crude Oil Production for the United States," by William Manzer, Western Wyoming College.
"What are the Effects (If Any) of a Mathematics Placement Exam?" by Duane Porter, University of Wyoming.
"A Moment on Moments," by Aubrey Owen, Community College of Denver.
"A Technique for Directly Evaluating the Mean Time to First Failure for Nonreparable Electronic Systems with Active Redundency," by John Garstka, Air Force Academy.
"Mathematical Learning Theory," by Edward De Francia, Fort Lewis College.
"Money-Math," by Ben Manvel, Colorado State University.
"Extrema in Polar Coordinates," by Aubrey Owen, Community College of Denver.
"Tackling a Ticklish Type of Tic-Tac-Toe (Or the Case of the [Almost] Total Tactics)," by Ira Rosenholtz, University of Wyoming.
"Computer-Generated Insights into Number Theory Results," by Robert Fisk, Colorado School of Mines and Technology.
"Integration: Why You Can and Why You Can't," by Rick Miranda, Colorado State University.
"Factoring Messy Trinomials," by Carl Kerns, Mesa College.
"New-Wave Cryptography," by Richard Games, Colorado State University.
"Service Courses: How the Engineers View What Mathematics Departments Provide," by David Ballew, South Dakota School of Mines and Technology.
"Zeros and Factors of Polynomials with Positive Coefficients," by William Briggs, University of Colorado.
"The Van Meegeren Art Forgeries," by James Coler, University of Colorado.

Kentucky Section

The spring meeting of the Kentucky Section was held at Bellarmine College on April 8-9, 1983. Eighty-five people were in attendance.

Invited Addresses:

"Why Mathematics is Useful," by Maynard Thompson, Indiana University.
"Applications of Mathematics in Bio-Medicine," by Maynard Thompson, Indiana University.
"Factorings in Graphs," by Gary Chartrand, Western Michigan University.

Invited Lectures:

"The Reconstruction Problem," by Don Greenwell, Eastern Kentucky University.
"Ramsey Theory," by Glenn Powers, Western Kentucky University.
"Directed Graphs," by Michael Jacobson, University of Louisville.
"Functional Obscurities of a Familiar Formula," by J.B. Barksdale, Western Kentucky University.
"Polya's Picture of Complex Functions," by Bart Braden, Northern Kentucky University.
"Subjective Probability and the Concept of Coherence," by Walter Ferbes, Bellarmine College.
"Microcomputer Courseware for Beginning Calculus," by Bill Shoalf, Murray State University.
"Number Theory and Public Key Cryptography," by Patrick Costello, Eastern Kentucky University.

Allegheny Mountain Section

The Allegheny Mountain Section held its annual spring meeting at Indiana University of Pennsylvania in Indiana on April 29-30, 1983.

Invited Lectures:

"Curves of the Calculus: History and Application," by V.F. Rickey, Bowling Green University.
"CUPM's New Mathematical Sciences Curriculum Recommendations," by Alan Tucker, SUNY at Stony Brook.
"Mathematics for a Computer Science Curriculum," by William Scherlis, Carnegie-Mellon University.
"Coalitional Games and Various Applications," by William Lucas, Cornell University.
"The Mathematical Sciences Curriculum K-12: What Is Still Fundamental and What Is Not," by Marcia Sward, Associate Director, MAA.

Panel Discussion:

"Putnam Exam: Ideas on Offering It and Taking It," Ron Harrell (Moderator), Allegheny College.

Short Presentations:

"The Next Computer Science Course After Programming," by David Brown, Bethany College.
"Fibonacci Numbers and Finding Extrema," by John Atkins, West Virginia University.
"Applications of Computer Graphics in Mathematics Instruction," by Roy Myers, Penn State University at New Kensington.

Student Presentations:

"Variations on Conway's Game of Life," by Tim Price, Allegheny College.
"What Can We Say About Duffian Numbers?" by Rosemary Zbiek, Indiana University of Pennsylvania.
"Tiling the Plane with Convex Polygons," by Kim Kratz, Allegheny College.
"Amicable Numbers," by Mark Woodard, Indiana University of Pennsylvania.

Maryland-District of Columbia-Virginia Section

The spring meeting of the Maryland-District of Columbia-Virginia Section was held April 16, 1983 at Thomas Nelson Community College in Hampton, Virginia.

Invited Address:

"Hardware, Software, and Mathematics," by Jack Beidler, University of Scranton.

Short Presentations:

"Almost Isosceles Pythagorean Triples," by William Wardlaw, U.S. Naval Academy.
"Environmental Mathematics," by Ben Fusaro, Salisbury State College.
"Constructions in Non-Euclidean Geometry," by George Ivey, Gallaudet College.
"What is a Sociology of Mathematics?" by J. Fang, Old Dominion University.
"Wedderburn's Finite Division Algebra Theorem: An Historical Perspective," by Karen Parshall, Sweet Briar College.
"Rudiments of Computational Support Distribution Test Space Topology," by Steven Mallis, Eastern Shore Community College.
"Adjacency Games/Squares," by Craig Bailey, U.S. Naval Academy.
"Adjacency Games/Hexagons," by Mark Kidwell, U.S. Naval Academy.
"License Plates and Phone Numbers--Statistics Classroom Learning Tools," by Edwin Landauer, General Physics Corporation.
"Tracking Algorithms Used with Search Radars," by Paul Ilacqua, J.S. Lee Associates.
"Computing Solutions to the Billiard Ball Problem," by Brian Shelburne, Sweet Briar College.

the following theorem is immediate.

THEOREM 4. *Let* $L_{2n}(y) = \sum_{k=0}^{2n} a_k y^{(k)}(x)$ *where* $a_k \in R$, $k = 0, 1, \ldots, 2n$ *and* $a_{2n} \neq 0$. *Then* $f(x) L_{2n}(y)$ *is formally self adjoint if and only if*:

(i) $$f(x) = e^{a_{2n-1}x/na_{2n}}$$

and

(ii) *for* $n \geqslant 2$,

$$a_{2k-1} = \sum_{j=0}^{n-k} \binom{2k+2j}{2k-1} \frac{(2^{2j+2}-1) B_{2j+2} a_{2k+2j} a_{2n-1}^{2j+1}}{(j+1) n^{2j+1} a_{2n}^{2j+1}}, \quad k = 1, 2, \ldots, (n-1).$$

3. Examples.

(1) $$L(y) = y^{(4)} - 5y^{(3)} + 5y'' + 5y' - 6y.$$

By Theorem 4, the symmetry factor is necessarily $f(x) = e^{(-5x)/2}$. However, condition (ii) of Theorem 4 is not satisfied so $L(y)$ cannot be made formally self adjoint.

(2) $$L(y) = y^{(4)} - 2y^{(3)} + y' + xy.$$

According to Theorem 3, the symmetry factor is necessarily $f(x) = e^{-x}$. It is easy to check to see that e^{-x} also satisfies condition (ii) of Theorem 3 so $e^{-x}L(y)$ is formally self adjoint. Note that

$$e^{-x}L(y) = (e^{-x}y'')'' - (e^{-x}y')' + xe^{-x}y.$$

(3) $$L(y) = x^2 y^{(4)} - (2x^2 - 4x) y^{(3)} + (x^2 - (2R+6)x) y'' + ((2R+2)x - 2R) y'.$$

Using Theorem 3, it follows that e^{-x} is a symmetry factor. Again, observe that

$$e^{-x}L(y) = (x^2 e^{-x} y'')'' - (([2R+2]x + 2) e^{-x} y')'.$$

(4) $$L(y) = y^{(6)} - 5y^{(5)} + 2y^{(3)} - y'' + 6y' + 3y.$$

The symmetry factor must necessarily be $f(x) = e^{-5x/3}$. However, condition (ii) of Theorem 4 is not satisfied so $L(y)$ cannot be made formally self adjoint.

Acknowledgements. The author would like to thank Professors H. L. Krall and A. M. Krall for their helpful comments and suggestions.

References

1. Coddington and Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.
2. R. H. Cole, Theory of Ordinary Differential Equations, Appleton-Century-Crofts, New York, 1968.
3. E. L. Ince, Ordinary Differential Equations, Dover, New York, 1951.
4. H. L. Krall, Self-Adjoint Differential Expressions, this MONTHLY, vol. 67, 9 (1960) 876–878.

EULER'S INTEGRALS

H. HARUKI
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

S. HARUKI
Department of Applied Mathematics, Okayama University of Science, 1-1 Ridai-cho, Okayama, 700, Japan

The aim of this note is to show an efficient method of evaluating the definite integral

(1) $$f(a) = \int_0^1 \frac{\log(x^2 - 2x\cos a + 1)}{x} dx \quad (0 \leqslant a \leqslant 2\pi).$$

Then the following three well-known (e.g., [**1**]) Euler's integrals can be evaluated immediately by applying our result:

$$\int_0^1 \frac{\log(1-x)}{x}\,dx = -\frac{1}{6}\pi^2, \tag{2}$$

$$\int_0^1 \frac{\log(1+x)}{x}\,dx = \frac{1}{12}\pi^2, \tag{3}$$

$$\int_0^1 \frac{\log(1+x^2)}{x}\,dx = \frac{1}{24}\pi^2. \tag{4}$$

Our method is rather unusual. We evaluate (1) by using solutions of a certain functional equation for a single variable and without using any of (2), (3), and (4).

THEOREM. *If f is defined by* (1), *then*

$$f(a) = -(a-\pi)^2/2 + \pi^2/6. \tag{5}$$

Proof. Apply the identity $(x^2 - 2x\cos a + 1)(x^2 + 2x\cos a + 1) = x^4 - 2x^2\cos 2a + 1$ to obtain

$$\begin{aligned} f(a/2) &+ f(\pi - a/2) \\ &= \int_0^1 \{[\log(x^2 - 2x\cos(a/2) + 1)(x^2 + 2x\cos(a/2) + 1)]/x\}\,dx \\ &= \int_0^1 \{[\log(x^4 - 2x^2\cos a + 1)]/x\}\,dx, \end{aligned}$$

which, with the transformation $x = \sqrt{t}$, implies

$$(1/2)\int_0^1 \{[\log(t^2 - 2t\cos a + 1)]/t\}\,dt = (1/2)f(a).$$

Hence f defined by (1) satisfies the functional equation

$$f(a/2) + f(\pi - a/2) = f(a)/2 \tag{6}$$

for all a in the closed interval $[0, 2\pi]$.

Since, by a basic theorem on differentiation under the integral sign, f is twice continuously differentiable, we differentiate (6) twice with respect to a to obtain

$$f''(a/2) + f''(\pi - a/2) = 2f''(a). \tag{7}$$

Further, f'' is a continuous function in the closed interval $[0, 2\pi]$. Hence $f''(a)$ has a maximum value M in $[0, 2\pi]$ and a minimum value m in $[0, 2\pi]$. Suppose that $f''(a_0) = M$ for a fixed $a_0 \in [0, 2\pi]$. Then by setting $a = a_0$ in (7) we have

$$f''(a_0/2) + f''(\pi - a_0/2) = 2f''(a_0) = 2M. \tag{8}$$

However, both $a_0/2$ and $\pi - a_0/2 \in [0, 2\pi]$, and

$$f''(a_0/2) \leqslant M, \quad f''(\pi - a_0/2) \leqslant M. \tag{9}$$

Hence we have

$$f''(a_0/2) + f''(\pi - a_0/2) \leqslant 2M. \tag{10}$$

So, by (8), (9), and (10) we must have $f''(a_0/2) = M$. By iteration $f''(a_0) = f''(a_0/2^n) = M$ and by continuity it follows that

$$\lim_{n\to\infty} f''(a_0/2^n) = f''(0) = M.$$

Similarly, we see that $f''(0) = m$. Thus $M = m$. Therefore, $f''(a)$ is a constant function in $[0, 2\pi]$. Let $f''(a) = \alpha$. Then f is given by the quadratic polynomial

$$f(a) = \alpha a^2/2 + \beta a + \gamma. \tag{11}$$

Our final step is to determine constants α, β, and γ explicitly. Substitute (11) in (6) and then equate the coefficients of the term a and the constant term to obtain

$$-\pi\alpha/2 = \beta/2, \quad \pi^2\alpha/2 + \beta\pi + 2\gamma = \gamma/2. \tag{12}$$

Notice that the coefficient of the term a^2 vanishes. By (1) and by a basic theorem on differentiation under the integral sign we obtain $f'(\pi/2) = \pi/2$, which, with (11), implies

$$\pi\alpha/2 + \beta = \pi/2. \tag{13}$$

It follows from (12) and (13) that $\alpha = -1$, $\beta = \pi$, and $\gamma = -\pi^2/3$. Thus, (11) yields (5). This completes the proof of Theorem.

The values of the Euler integrals (2), (3) and (4) now follow from (1) and (5) by setting a equal to 0, π, and $\pi/2$ respectively.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada Grant Nr. A-4012.

Reference

1. R. Ayoub, Euler and the zeta function, this MONTHLY, 81 (1974) 1067–1086.

A HOMOLOGY VERSION OF THE BORSUK-ULAM THEOREM

JAMES W. WALKER
Department of Mathematics, California Institute of Technology, Pasadena, CA 91125

An **involution** on a topological space X is a continuous map from X to X which is its own inverse. For example the antipodal map, which maps a point to the opposite end of the diameter on which it lies, is an involution on the n-sphere S^n.

Suppose X and Y are spaces equipped with involutions a and b, respectively. A map f from X to Y is **equivariant** if it respects the involutions, i.e., $b \circ f = f \circ a$.

One formulation of the Borsuk-Ulam theorem is that if m is greater than n, then there is no map from S^m to S^n which is equivariant with respect to the antipodal map. Many sources, for example [**1**, § 7.2], include proofs of the Borsuk-Ulam theorem, as well as applications such as the "ham sandwich theorem." We will use singular homology theory to prove a somewhat stronger theorem.

Our stronger theorem shows that the existence of any equivariant maps to S^n from any space X with an involution forces the existence of very special homology classes for X, so special that X could not be a sphere of dimension greater than n.

A few words about terminology: An **elementary 0-chain** is a singular 0-simplex with coefficient 1; loosely speaking, it's just a single point. We will use reduced homology, which essentially means that we consider the empty set to be a singular simplex of dimension -1, which is the boundary of every 0-simplex. It follows that $\tilde{H}_{-1}(X)$ vanishes unless X is empty, and $\tilde{H}_0(X)$ vanishes if X is path connected. Recall that each continuous map f induces a chain map $f_{\#}$, defined by composing f with singular simplices. In turn, such a chain map $f_{\#}$ induces a homology homomorphism f_*.

THEOREM. *Suppose X is a space with involution v, and $g\colon X \to S^n$ is an equivariant map. Then there exists an integer $j \leqslant n$, and a homology class β of $\tilde{H}_j(X; \mathbb{Z}/2)$ such that β is nonzero and $v_*(\beta) = \beta$. Furthermore, if no such β exists for j less than n, then β can be chosen such that $g_*(\beta)$ is the nonzero element of $\tilde{H}_n(S^n; \mathbb{Z}/2)$.*

homologous to either zero or θh_n. In either case, when we apply θ, we find that $\theta h_n - g_{\#}\theta c_n$ is homologous to zero. That is, θh_n and $g_{\#}\theta c_n$ belong to the same homology class. Note that θc_n is a cycle, because $\partial\theta c_n = \theta\partial c_n = \theta\theta c_{n-1} = 0$. Therefore, if β is the homology class of θc_n, then $g_*(\beta)$ is the nonzero element of $\tilde{H}_n(S^n; \mathbb{Z}/2)$. It follows that β is nonzero. Finally, the fact that $\theta\theta c_n = 0$ means that $v_{\#}\theta c_n = \theta c_n$, so $v_*(\beta) = \beta$.

References

1. M. K. Agoston, Algebraic Topology, Marcel Dekker, New York, 1976.
2. D. G. Bourgin, Modern Algebraic Topology, Macmillan, New York, 1963.

ON THE NUMBER OF MULTIPLICATIVE PARTITIONS

JOHN F. HUGHES
Department of Mathematics, Bryn Mawr College, Bryn Mawr, PA 19010

J. O. SHALLIT
Department of Mathematics, University of California, Berkeley, CA 94720

I. A Number-Theoretic Function. In this note we show that if $f(n)$ is the number of essentially different factorizations of n, then

$$f(n) \leqslant 2n^{\sqrt{2}}.$$

In considering numbers that have exactly k divisors, one is led to examine this function $f(n)$, the number of ways to write n as the product of integers $\geqslant 2$, where we consider factorizations that differ only in the order of the factors to be the same. We call these representations of n **multiplicative partitions**. For example, $f(12) = 4$, since

$$12 = 6 \cdot 2 = 4 \cdot 3 = 3 \cdot 2 \cdot 2$$

are the four multiplicative partitions of 12. From these four representations, we can conclude that a number has exactly 12 divisors if and only if its prime factorization is one of the following:

$$p^{11}, p^5q, p^3q^2, p^2qr.$$

This follows from the expression for $\tau(n)$, the number of divisors of $n = p_1^{a_1}p_2^{a_2} \cdots p_k^{a_k}$.

$$\tau(n) = \prod_{j=1}^{k} (1 + a_j).$$

For example, see [1].

The behavior of $f(n)$ is quite erratic, and apparently has not been previously studied in this form. We observe that if q is prime, then $f(q^k) = p(k)$, the number of additive partitions of k. Also, if $q_1, q_2, \ldots, q_k$ are distinct primes, then $f(q_1q_2 \cdots q_k) = B(k)$, the kth Bell number. See [2].

More generally, $f(q_1^{a_1} \cdots q_k^{a_k})$ is the number of additive partitions of the "multi-partite number" $(a_1, a_2, \ldots, a_k)$, where addition is defined component-wise. See [3] for further details.

We will show that

$$f(n) \leqslant 2n^{\sqrt{2}}. \tag{1}$$

For a table of $f(n)$ for $1 \leqslant n \leqslant 100$, see the Appendix.

II. Proof of the Main Result. To prove (1) we first define an auxiliary function:

$g(m, n) =$ the number of multiplicative partitions of n with all elements $\leqslant m$.

Clearly $f(n) = g(n, n)$. We have the following

THEOREM 1.

$$g(m,n) = \sum_{\substack{d|n \\ d\leqslant m}} g(d, n/d). \tag{2}$$

Proof. We define $g(m,1) = 1$ and $g(1,n) = 0$ for $n \neq 1$. Let $n = a_1 a_2 \cdots a_k$ be a multiplicative partition of n with all factors $\leqslant m$. Then we may assume the factors are arranged in decreasing order, so a_1 is the largest factor in the product. The number of ways to choose $a_2 \cdots a_k$ is therefore $g(a_1, n/a_1)$. But a_1 was unspecified, and therefore could be any divisor d of n such that $d \leqslant m$. Summing over all such d gives the result. □

From Theorem 1 we can obtain a simple estimate for $g(m,n)$.

THEOREM 2.

$$g(m,n) \leqslant mn.$$

Proof. The theorem is clearly true for $m = 1$ or $n = 1$. We will show it is true by induction on the product mn. Assume true for all m, n such that $mn < MN$, where $M \geqslant 2$. Then from Theorem 1 we have

$$g(M,N) = \sum_{\substack{d|N \\ d\leqslant M}} g(d, N/d).$$

Since $d \cdot N/d = N < MN$, we may apply the induction hypothesis to the terms inside the summation. We find

$$\begin{aligned} g(M,N) &\leqslant \sum_{\substack{d|N \\ d\leqslant M}} d \cdot N/d \\ &\leqslant \sum_{d\leqslant M} N \\ &= MN, \end{aligned}$$

and the theorem is true by induction. □

Theorem 2 gives the estimate $f(n) = g(n,n) \leqslant n^2$. It is possible to improve this estimate, which we do in a moment. First we need three easy lemmas.

LEMMA 3.

$$g(a,b) \leqslant g(b,b).$$

Proof. This follows immediately, since if $a \geqslant b$, we have strict equality, while if $a < b$, we have summing over fewer terms of equation (2). □

LEMMA 4. *Let* $0 < c < 1$. *Then*

$$f(n) \leqslant g(n^c, n) + \sum_{d=1}^{n^{1-c}} f(d).$$

Proof.

$$\begin{aligned} f(n) = g(n,n) &= \sum_{d|n} g(d, n/d) \\ &= \sum_{\substack{d|n \\ d\leqslant n^c}} g(d, n/d) + \sum_{\substack{d|n \\ d> n^c}} g(d, n/d) \end{aligned}$$

$$\leqslant g(n^c, n) + \sum_{\substack{d|n \\ d>n^c}} g(n/d, n/d) \text{ (by Theorem 1 and Lemma 3)}$$

$$= g(n^c, n) + \sum_{\substack{d|n \\ d<n^{1-c}}} g(d, d)$$

$$\leqslant g(n^c, n) + \sum_{d=1}^{n^{1-c}} f(d),$$

which is the desired result. □

LEMMA 5. *Let $a \geqslant 0$. Then*

$$\sum_{d=1}^{k} d^a \leqslant \frac{k^{a+1}}{a+1} + k^a.$$

Proof. This is easily proved by comparison with the integral $\int_1^k t^a \, dt$.

We are now in a position to prove our main result.

THEOREM 6.

$$f(n) \leqslant 2n^{\sqrt{2}}.$$

Proof. The table in the Appendix shows the theorem is true for $n \leqslant 69$. We will prove the theorem by induction on n. Assume $f(d) \leqslant kd^{c+1}$ for $d < n$, where $n \geqslant 70$ and c and k are constants to be specified later. Then from Lemma 4 we have

$$f(n) \leqslant g(n^c, n) + \sum_{d=1}^{n^{1-c}} f(d)$$

$$\leqslant n^{c+1} + \sum_{d=1}^{n^{1-c}} f(d) \text{ (by Theorem 2)}$$

$$\leqslant n^{c+1} + k \sum_{d=1}^{n^{1-c}} d^{c+1} \text{ (by induction)}$$

$$\leqslant n^{c+1} + k\left(\frac{(n^{1-c})^{c+2}}{c+2} + (n^{1-c})^{c+1}\right) \text{ (by Lemma 5)}.$$

Now put $k = 2$ and $c = \sqrt{2} - 1$ to get

$$f(n) \leqslant n^{\sqrt{2}} + \frac{2}{\sqrt{2}+1} n^{\sqrt{2}} + 2n^{2(\sqrt{2}-1)}$$

$$\leqslant 2n^{\sqrt{2}}$$

since $2/\sqrt{2} + 1 < 5/6$ and $2n^{2(\sqrt{2}-1)} \leqslant 1/6n^{\sqrt{2}}$ for $n \geqslant 70$.

Our theorem is now proved by induction. □

III. Two Conjectures. Numerical evidence seems to indicate that the exponent $\sqrt{2}$ in Theorem 6 is too large. We make two conjectures; the second is more doubtful.

CONJECTURE 1.

$$f(n) \leqslant n.$$

CONJECTURE 2.

$$f(n) \leqslant \frac{n}{\log n} \text{ for } n \neq 144.$$

Both these conjectures have been verified by computer for $n \leqslant 10{,}000$.

Appendix

n	$f(n)$	n	$f(n)$	n	$f(n)$	n	$f(n)$
1	1	26	2	51	2	76	4
2	1	27	3	52	4	77	2
3	1	28	4	53	1	78	5
4	2	29	1	54	7	79	1
5	1	30	5	55	2	80	12
6	2	31	1	56	7	81	5
7	1	32	7	57	2	82	2
8	3	33	2	58	2	83	1
9	2	34	2	59	1	84	11
10	2	35	2	60	11	85	2
11	1	36	9	61	1	86	2
12	4	37	1	62	2	87	2
13	1	38	2	63	4	88	7
14	2	39	2	64	11	89	1
15	2	40	7	65	2	90	11
16	5	41	1	66	5	91	2
17	1	42	5	67	1	92	4
18	4	43	1	68	4	93	2
19	1	44	4	69	2	94	2
20	4	45	4	70	5	95	2
21	2	46	2	71	1	96	19
22	2	47	1	72	16	97	1
23	1	48	12	73	1	98	4
24	7	49	2	74	2	99	4
25	2	50	4	75	4	100	9

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford, Clarendon Press, 1971, p. 239.

2. G. T. Williams, Numbers generated by the function $e^{e^{x}-1}$, this MONTHLY, 52 (1945) 323–327.

3. George Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications 2, Gian-Carlo Rota, Editor, Addison-Wesley, Reading, Mass. 1976.

ANSWERS TO PHOTOS ON PAGE 437

No, they are partial. They are two of the most famous partial differential equators in the world. Top: Lars Hörmander of Lund; bottom: Olga Ladyženskaja of Leningrad.

COMMENTS AND COMPLEMENTS

DEBORAH AND FRANKLIN TEPPER HAIMO
Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121

Members of our alert readership periodically communicate to us comments about various papers published in the Notes sections. They identify errors, supply added references, provide alternative proofs, introduce historical supplements, point out duplications in the literature, and bring various other items of interest to our attention. Having relinquished our editorial responsibilities, we take this opportunity to share correspondence of general interest that has accumulated over recent months.

Combinatorics. John R. Hunt of the University of Wisconsin in Menomonie points out that on page 630 of the paper, "Infinite products for kth roots," by N. J. Fine (this MONTHLY, 84 (1977) 629–630), the number appearing in equation (10) as 140541, found by the recursive relation $q_{n+1} = q_n^3 + 3q_n^2 - 3$, should be 140451.

David Zeitlin of Minneapolis, Minnesota, writes that his abstract in the *Notices* of the AMS, 26 (1969) A-527, entitled "$T(N)$, the number of incongruent triangles with integer sides and perimeter N, is a solution of a homogeneous, linear difference equation of minimal order nine (9)," contains the result that $T(N)$ satisfies the equation

$$\begin{aligned} T(N+9) = T(N+7) + T(N+6) + T(N+5) \\ -T(N+4) - T(N+3) - T(N+2) + T(N), \end{aligned}$$

for $N = 0, 1, 2, \ldots$. A generating function for $T(N)$ is given by

$$(x^2-1)^2(x^2+1)(1-x^3)\sum_{N=0}^{\infty} T(N)x^N = x^3.$$

His comments are in reference to the statement in the paper, "Triangles with integer sides," by J. H. Jordan, Ray Walch, and R. J. Wisner (this MONTHLY, 86 (1979) 686–689), that $T(N)$ "seems confusing if not intractable." A further remark is that, on page 188 of L. Dickson's "History of the Theory of Numbers" (Chelsea, N.Y., 1952), a 1908 paper by R. W. D. Christie is cited that deals with a related problem of dissimilar rational triangles of equal perimeter.

In the paper, "Iterated binomial coefficients," by Solomon W. Golomb (this MONTHLY, 87 (1980) 719–727), the author states that "no simple reduction formulae have yet been found for the most general case of $\binom{\binom{n}{b}}{a}$." According to David Zeitlin, this is not the case since, in the *Abstracts* of the AMS, 2 (1981) 290, Zeitlin's paper, "General reduction formulas for iterated binomial coefficients," solves the problem. When Golomb's attention was called to the Zeitlin abstract, he commented that he would hesitate to label Zeitlin's results as "the solution." He notes that there are many reduction formulas and Zeitlin's fail to be simple in the sense that the coefficient C_j and D_j are themselves complicated summations involving both two-tiered and three-tiered binomial coefficients. He acknowledges, however, that he has not seen anything simpler.

V. N. Murty of the Pennsylvania State University cites G. Chrystal's "Textbook of Algebra" (Chelsea, vol. 2, 7th ed., 1964), where the results appearing in Gabriel Klambauer's paper, "Summation of series" (this MONTHLY, 87 (1980) 128–130), are also given. In particular, the author's examples illustrating technique I are included in Chrystal's text on page 336, and the technique, known as early as 1870, is due to Laisant. Further, technique II of the paper is given in Chrystal's Chapter XXXI, page 419, section 12. Murty comments that Chrystal has an excellent collection of methods of summing infinite series using elementary methods.

Number Theory. B. M. Stewart of Okemos, Michigan, in commenting on C. M. Cordes' paper,

"Permutations mod m in the form x^n" (this MONTHLY, 83 (1976) 32–33), suggests the alternate formulation of the problem as follows: "Find the smallest $s > 0$ so that each of the congruences $x^s = a \bmod m$ for $a = 0, 1, \ldots, m-1$ has exactly one solution." He remarks that the existence of s is contingent upon m being square free. In that case, if $m = p_1 \cdots p_k$ for an ascending set of primes p_i, then the least s must be the smallest prime q such that $(q, (p_1 - 1) \cdots (p_k - 1) = 1$.

David Zeitlin points out Moshe Lotan's "A problem in difference sets" (this MONTHLY, 56 (1949) 535–541) as a useful reference for Robert Miller's "A game with n numbers" (this MONTHLY, 85 (1978) 183–185), and cites, in addition, the Italian journal, *Periodiche di Mathematiche* (17 (1937) 25–30).

Referring to a comment attributed to Ray Davis and noted on page 837 of the "Comments and Complements" article (this MONTHLY, 86 (1979)), Joseph Berlau of Hartsdale, New York, questioned, and Ralph Boas of Northwestern University clarified the meaning of 300.999508 as an irrational number that starts with these digits.

Algebra. While commenting that Phyllis Joan Cassidy's note, "Products of commutators are not always commutators: an example" (this MONTHLY, 86 (1979) 772) "gives the most elegant proof [he] know[s], even for the case $n = 1$, of the following result: There exists a group G such that, for all n, not every element of the commutator subgroup can be written as a product of n commutators," Roger Lyndon of the University of Michigan is concerned that "the very modest list of references given might mislead some readers in that it does not indicate the rather considerable amount of work that has been published on this and related problems." He cites two examples. In his book with Paul E. Schupp, "Combinatorial group theory" (*Ergebnisse der Math.*, 89, Springer Verlag, Berlin, Heidelberg, New York, 1977), on page 55, it is stated that a certain free group G has the property mentioned above; and, in the paper by Basil Jordan, Robert M. Guralnick, and Michael D. Miller "On cyclic commutator subgroups" (*Aequationes Math.*, 17 (1978) 112–113), it is stated that Guralnick has constructed a group G of order 96 in which not every element of the commutator subgroup is a commutator, and that this is the smallest order for such a group. Asserting that the problem considered can be put under the general heading of Equations in Groups, Lyndon points out that some notes on this subject include those of J. Mycielski "Can one solve equations in groups?" (this MONTHLY, 84 (1977) 723–726), and "Equations unsolvable in $GL_2(\mathbb{C})$ and related problems" (this MONTHLY, 85 (1978) 263–265), and his own paper "Equations in groups," with an extensive bibliography, especially on the question of products of commutators, to appear in *Boletin da Soc. Brasileira da Math*, 11 (1980).

Raphael M. Robinson of the University of California, Berkeley, informs us that his research problem, "Solutions of an equation in abelian groups" (this MONTHLY, 86 (1979) 690), has generated five solutions. One, by Sidney C. Garrison, Martin R. Perlet, and Stephen M. Gagola, all of Texas A and M University (this MONTHLY, 88 (1981) 195–196), uses dual groups, and explicitly reaches the conclusion

$$\sum_{i=1}^{n} \frac{1}{|g_i|} > 1, n > 1,$$

for the equation $\prod_{i=1}^{n} (1 - g_i) = 0$ in abelian groups, where $|g_i|$ is the order of g_i. Other similar solutions were submitted by Alfred W. Hales of the University of California, Los Angeles, Masao Kiyota of the University of Tokyo, and Geoffrey R. Robinson of Coventry, England. Ki Haug Kim and Fred W. Roush of Alabama State University proposed a solution referring to group representations.

Geometry. Leon Gerber of St. John's University in Jamaica, New York, provides historical perspective for some published notes on geometry. As references for Dixon Jones's notes, "Quadrangles, butterflies, Pascal hexagons, and projective fixed points" (this MONTHLY, 87 (1980) 197–200) he cites "Poncelet's theorems, 'Traité des propriétés projectives des figures'" (Paris, 1822, article 513), and page 147 of Cremona's "Elements of Projective Geometry" (1913).

Historical sources for the 3-dimensional version of E. Snapper's "Affine generalization of the

Euler line (this MONTHLY, (1981) 196–198), according to Leon Gerber, include *Educational Times Reprints*, 19 (1873) 38–39 and N. A. Court's "Modern Pure Solid Geometry," Chelsea, 2nd ed., 1964, as Ex. 11 on page 121. Further, an n-dimensional version is on page 85 of geometry notes Gerber himself published in 1975 for use in a seminar at St. John's University.

Commenting on Walter Stomquist's paper, "How to cut a cake fairly" (this MONTHLY, 87 (1980) 640–644), A. Keith Austin of the University of Sheffield, England, writes that "the algorithm on page 641 fails in the case where the knife of the player who shouted 'cut' is not the center knife. In this case, the player who shouted is not satisfied about the piece received by one of the other players, as he thinks it is too large." On being apprized of this observation, Walter Stomquist writes that Dr. Austin is quite right in this respect: "If a player shouts 'cut' on the assumption that the cake will be cut by the sword and his own knife, then he may feel slighted if the cut is made by another player's knife instead. (He will get 1/3 of the cake, but as [was pointed out], he will regard another piece as larger.) I should have made clear that, when deciding when to say 'cut,' *each* player should watch the sword and the *middle* knife. That is actually why the knives are manipulated so publicly. In this way, the player can delay shouting until he will be satisfied by the left piece as it actually will be cut."

Analysis. A note from John P. Hoyt of Lancaster, Pennsylvania, expresses surprise that the editorial staff of the MONTHLY permitted the deletion of the "De" from De Moivre's name in J. van Yzeren's note "Moivre's and Fresnel's integrals by simple integration" (this MONTHLY, 86 (1979) 691–693). "What is so 'difficult' about 'De Moivre'? 'De Moivre' it has been, 'De Moivre' let it be." He points out that even such a nonmathematical source as *The Encyclopedia Britannica* refers to "De Moivre." Noting that the "van" has been included in the author's name, he asks plaintively whether De Moivre is "to suffer the fate of 'Tchebycheff' and L'Hopital' [sic]."

Hans Samelson writes with some embarrassment that it has come to his attention that his proof for Rolle's theorem appearing in his paper, "On Rolle's theorem" (this MONTHLY, 86 (1979) 486), is identical with that given by R. J. Easton and S. G. Wayment in their paper, "The sliding interval technique" (this MONTHLY, 75 (1968) 886–888). He adds, further, that these authors gave a generalization of the theorem to E^n (this MONTHLY, 77 (1970) 170–172) also. These facts escaped not only the author but three knowledgeable referees and the editors as well.

Providing an addendum to his note, "An ultimate proof of Rolle's theorem," Alexander Abian states that "lines 16 to 13 from the bottom of page 485 (this MONTHLY, 86 (1979)) should be replaced by the following, which extends almost verbatim [the] proof of $c \in (a_0, b_0)$ to $c \in (a_i, b_i)$ for every $i \in \omega$:... [For,] from (1) and (2) we have $g(m_{i+k}) \geqslant \max\{g(a_i), g(b_i)\}$ for every k. Now, if $g(m_{i+k}) > \max\{g(a_i), g(b_i)\}$ for some k, then from (6) and (9) it follows that $g(c) > \max\{g(a_i), g(b_i)\}$, which implies $a_i \neq c \neq b_i$ and therefore $c \in (a_i, b_i)$. On the other hand, if $g(m_{i+k}) = \max\{g(a_i), g(b_i)\}$ for every k, then from (2) and (8) it follows that $c = m_i$, which again implies $c \in (a_i, b_i)$."

David Zeitlin cites page 12 of I. J. Schwatt's "An Introduction to the Operation with Series" (second edition (reprint of the 1924 first edition), Chelsea, New York), where a related formula is included to that contained in Steven Roman's paper, "The formula of Faa Di Bruno" (this MONTHLY, 87 (1980) 805–809). The companion result is an expansion for $D_t^n g(f(t))$ which, according to Zeitlin, is easier to use for computational purposes than Faa Di Bruno's formula.

Colin C. Graham of Northwestern University, commenting on A. M. Russell's paper, "A Commutative Banach algebra of functions of bounded variation," points out that Nina Bau in a paper in the *Mathematische Annalen* in 1930 knew that BV, and hence BV^*, are closed under multiplication. Though she did not state the norm inequality used by Russell, that inequality has been known at least since 1957, when it appeared in a paper by Hewitt and Zuckerman in the *Pacific Journal of Mathematics* (7 (1957) 913–941).

Robert S. Doran of Texas Christian University writes that not only is W. E. Pfaffenberger's

Theorem 1 in his paper, "A converse to a completeness theorem" (this MONTHLY, 87 (1980) 216), well known, as the author acknowledges, but that Theorem 2 is also well known. He cites Seymour Goldberg's book, *Unbounded Linear Operators* (McGraw-Hill, 1966), where, in Corollary I.5.8, page 21, a simpler proof is given, and Sterling Berberian's "Introduction to Hilbert Space," (Oxford, 1961), where on page 193 a short, elegant solution is given to the result stated as Exercise 12 on page 108.

The fact that Theorem 2 is well known was also pointed out by Chaitan P. Gupta of Northern Illinois University and by A. Venugopalan of Calicut, India.

Joel Brenner of Palo Alto, California, points out that appealing to the method that Adolf Cusmariu used in his note, "A proof of the arithmetic mean-geometric mean inequality" (this MONTHLY, 88 (1981) 192–194), one can prove that, for positive real numbers $x_1,\dots, x_n$, not all equal, the inequality

$$(n-1)(x_1+\cdots+x_n)^2 > 2n(x_1x_2+\cdots+x_{n-1}x_n)$$

holds.

Commenting on U. V. Satyanarayana's paper, "A note on Riemann-Stieltjes integrals" (this MONTHLY, 87 (1980) 477–478), Roy O. Davies, of the University of Leicester, writes that, "essentially the same proof was given for essentially the same result (Riemann integrals instead of Riemann-Stieltjes but the extension is trivial) in Mathematical Note 2970, 'A fundamental inequality for integrals,' by J. Sr.-C. L. Sinnadurai [*Mathematical Gazette*, 45 (1961) 235–236; corrigenda ibid 46 (1962) 143]." That observation also was made both by the latter author himself and by D. F. Scrimshaw of Brunel University. Max L. Weiss, of the University of California, Santa Barbara, provides the background information that Satyanarayana's interest in and proof of the theorem in question originated years ago as a consequence of "a loose end in one of my classes in a term paper concerning line integrals in connection with Cauchy's Theorem." He further notes that "Sinnadurai's proof is mathematically superior to Satyanarayana's. The former proves an interesting lemma about positive functions from which the result follows easily, while the latter involves the definition of the integral from the beginning."

On having the Sinnadurai paper called to his attention, Satyanarayana reached the similar conclusion that the former had a "more elegant" proof, but commented that "the method of my proof seems to be the most natural one any student (with proper preparation) would also come up with ... and the emphasis of my note is on pedagogy."

David Zeitlin writes that in his paper "An inequality for a class of polynomials" (*Fibonnacci Quarterly*, 16 (1978) 128–129, 146, 151), he generalizes equation (18) on page 29 of the paper "Inequalities and identities for sums and integrals" by M. S. Klamkin and D. J. Newman (this MONTHLY, 83 (1976) 26–30).

In commenting on J. M. Henle's note, "Functions with arbitrarily small periods" (this MONTHLY, 87 (1980) 816), Ralph Boas of Northwestern University provides yet another proof of Burstin's theorem that a Lebesgue-measurable function having arbitrarily small periods is constant a.e. He notes that if y is a period of a bounded, measurable function having dense periods, then

$$\int_0^x f(t+y)\,dt = \int_y^{x+y} f(t)\,dt = \int_0^x f(t)\,dt.$$

If, now,

$$F(x) = \int_0^x f(t)\,dt,$$

then F is a continuous solution of $F(x+y) = F(x) + F(y)$, for all x and a dense set of y's. Hence

$$F(x) = \lambda x = \int_0^x \lambda\,dt,$$

or

$$\int_0^x [\lambda - f(t)]\, dt \equiv 0.$$

It follows that

$$\int_a^b [\lambda - f(t)]\, dt = 0$$

for every (a, b), and $\lambda - f(t) = 0$ a.e. Henle's remark on seeing this proof was that it was an "interesting proof, though not self-contained."

Bruce R. Johnson of the University of Victoria, British Columbia, writes that he has just been alerted to the Lajos Takács paper, "On the method of inclusion and exclusion," in the *Journal of the American Statistical Association*, 62 (1967) 102–113. It contains, he points out, the same proof as that of his note, "An elementary proof of inclusion-exclusion formulas to real or complex-valued, finitely additive set functions." An extensive list of references on the subject is cited in the Takács paper.

Julian Hennefeld's question in his paper, "A nontopological proof of the uniform boundedness theorem" (this MONTHLY, 87 (1980) 217) about why Hausdorff's proof "fared so poorly in comparison to rival proofs" and his remark that "textbooks in analysis and functional analysis give only the Baire category proofs without any mention of an elementary proof in the text or even as an exercise" prompted Béla Szökefalvi-Nagy of Institutum Bolyaianum Universitatis, Hungary, to respond that this observation is not completely accurate. Essentially, the same nontopological proof appears, in section 31 of Riesz-Sz.-Nagy's "Functional Analysis" (Ungar, 1955), but relies directly on an 1897 paper by Osgood, except for obvious changes needed to avoid topology beyond the completeness property of the space. In section 87 of that text, Hausdorff's work also is referred to. Sz.-Nagy urges young mathematicians not to overlook classical authors like Osgood, for example.

Albert Wilansky of Lehigh University supplements the reference above with G. G. Lorentz's "Approximation of Functions" (Holt, Rinehart, Winston, 1966), where on page 95 a proof without category appears. He notes, further, that Hilbert-space proofs are included in S. S. Holland's note, "A Hilbert space proof of the Banach-Steinhaus theorem" (this MONTHLY, 76 (1969) 40–41) and in Stone and Tamarkin's paper, "Elementary proofs of some known theorems of the theory of complex euclidean spaces" (*Duke Mathematical Journal*, 3 (1937) 294–302). In his own book, "Modern Methods in Topological Vector Spaces" (McGraw-Hill, 1978), Wilansky uses the category argument to prove his Theorem 3-3-6 because it is brief and he wants to move on; because it motivates the idea of barrelled space; and because the theorem holds for all spaces of second category (not necessarily complete ones). The modern trend toward proving functional analysis results by elementary methods prompted Wilansky's colleague, G. A. Stengle, to remark, "That's all very well, but does it generalize?"

William C. Waterhouse of the Pennsylvania State University points out that the main theorem in H. Hueber's note, "On uniform continuity and compactness in metric spaces" (this MONTHLY, 88 (1981) 204–205), was proved earlier in M. Atsuji's paper, "Uniform continuity of continuous functions of metric spaces" (*Pacific J. Math.*, 8 (1958) 11–16), and that further results on spaces where all continuous functions are uniformly continuous can be found in J. Rainwater's "Spaces whose finest uniformity is metric" (*Pacific J. Math.*, 9 (1959) 567–570) and in his own note, "On UC spaces" (this MONTHLY, 72 (1965) 634–635).

Referring to the same paper, Angus E. Taylor writes that on page 204 Edwin Hewitt is given credit for answering a certain question in 1948 where, in actuality, the result dates to the thesis of Maurice Fréchet "who invented metric spaces and originated the term 'compact.' Fréchet called a set S compact if every infinite subset of S has a limit point (in the space, not necessarily in S). In his thesis, Fréchet showed that every closed and compact set in a metric space has the Borel property (that from every denumerable open covering of the set one can extract a finite open

covering). Then he proved a theorem that contains the result that if a metric space is not compact, one can define on it a real continuous function that is not bounded. The full theorem, on page 31 of his published thesis, is that a set S in a metric space is closed and compact if and only if every continuous real function on S is bounded and attains its least upper bound. In Fréchet's very first paper on abstract point-set topology, for an L-space (more general than a metric space), he proved that a continuous real function on a closed and compact set in such a space is necessarily bounded and attains its least upper bound (*Comptes Rendus*, 139 (1904) 848–850)."

Finally, Albert Wilansky of Lehigh University notes that, although his book, "Topology for Analysis" (Wiley, 1970), contains Hueber's results, he himself copied the results from articles of Mrowka, Waterhouse, and Atsuji, and that MONTHLY notes "are intended to bring such results to the attention of readers who do not consult the research literature."

When the observations above were made known to him, H. Hueber responded that, not only he, but several noted mathematicians who had read his paper before it was submitted were unaware of these results. That, of course, is true also of the two knowledgeable referees consulted by the editors. He provided the added information that his proof that (i) $\Rightarrow$ (ii) of Theorem 1 contains a misleading argument, which arose as a consequence of his attempt to find the shortest and most elegant formulation of the theorem. He points out that, where he wrote that "as k tends to infinity, $d(x_{n_k}, y_{n_k})$ tends to zero, and hence $(f(x_{n_k}) - f(y_{n_k}))/d(x_{n_k}, y_{n_k})$ tends to infinity," which is correct but misleading, it would have been better to write, "as k tends to infinity $d(x_{n_k}, y_{n_k})$ tends to zero whereas $f(x_{n_k}) - f(y_{n_k}) \to 1$."

David Zeitlin writes that his paper, "A derivation of the general solution for homogeneous, linear difference equations with constant coefficients" (this MONTHLY, 68 (1961) 369–370), treats, from a slightly different point of view, the same problem as that in Albert Nijenhuis' note, "Complete solutions of linear difference equations" (this MONTHLY, 87 (1980) 658–660). On being alerted to Zeitlin's result, which he had not been aware of, Nijenhuis' response was that "Zeitlin's theorem and its proof ... states an interesting relationship between difference equations and differential equations, but it does not explicitly address itself to a proof that [the solution given in my note] is, in fact, the general solution. No doubt such a proof is easily obtained, and Zeitlin can produce one at a moment's notice—quite possibly using some of the lines of computation in the paper."

B. M. Stewart of Michigan State University notes that he had anticipated several of the results in the article by J. C. Hintz (this MONTHLY, 87 (1980) 212–215) in a paper, "Sums of functions of digits," which appeared in the *Canadian Journal of Mathematics* (12 (1960) 374–389).

We have heard from Donald E. G. Malm of Oakland University that the algorithm for solving a linear diophantine equation given in Stanley Kertzner's note, "The linear Diophantine equation" (this MONTHLY, 88 (1981) 200–203), appears in a number of other sources including W. A. Blankinship's paper, "A new version of the Euclidean algorithm" (this MONTHLY, 70 (1963) 742–745), and Paul Schaefer's, "An algorithm for finding the G.C.D. of a finite set of integers" (*Delta*, 5 (1975) 19–27), as well as his own text, "A Computer Laboratory Manual for Number Theory" (student manual) (COMPress (1980) 89–95). In the latter, there is an extension of the algorithm to solve systems of simultaneous linear diophantine equations, together with programs in BASIC implementing the algorithm.

Stanley Kertzner wrote that he is chagrined to learn that, in checking only the text rather than the journal literature as well, he missed both the Schaefer and Blankinship papers. Since he had submitted his paper in December of 1979, however, the Malm Manual was not yet available. Subsequently, he has developed a sharpened procedure to solve linear integer matrix equations of the form $AX = B$ for all integer solutions, extending the algorithm that he had generalized to a system of diophantine equations.

That Blankinship's algorithm is not in popular use was decried by E. F. Assmus, Jr., of Lehigh University who noted that Daniel A. Marcus's "An alternative to Euclid's algorithm" (this MONTHLY, 88 (1981) 280–283), is far more complicated than the earlier published version.

Marcus, on the other hand, objects to the characterization of Blankinship's algorithm as "simpler," arguing that the two are significantly different, with the Blankinship algorithm having the advantage of applying to arbitrarily many integers rather than just two, of requiring no backtracking, and of relying on the familiar concept of row operations. His own algorithm, on the other hand, has the advantage of requiring that only the sequence of remainders be recorded and of producing the solutions by a single string of operations.

The publication of M. Hendy's note, "The retrial of the lower Slobbovian counterfeiters" (this MONTHLY, 87 (1980) 200–201), evoked for Juan Jorge Schäffer of Carnegie-Mellon University the memory that, when Problem E1096 appeared in this MONTHLY, 61 (1954) 46, he was a student in Zürich. P. Erdős, who happened to visit at that time, mentioned the problem to him, and they "soon found the fallacious 'solution' that had presumably been submitted in the meantime to the MONTHLY, and then discovered the fallacy, as well as the correct solution. He goes on to say that, when later that year he found that the fallacious "solution" had been published (this MONTHLY, 61 (1954) 472–473), he "wrote the editor of the Elementary Problems section to point out the error and provide a correct solution." When he received no acknowledgement, and the error was not corrected nor was a valid solution published, he did not follow up the matter. He encloses a copy of his 1954 letter "as a curiosity," having "nothing to complain about" and notes that "I shall be able to continue to boast that my Erdős number is 1, but that it is unlisted."

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

UNDERGRADUATE TRAINING FOR INDUSTRIAL CAREERS

ANN K. STEHNEY
Department of Mathematics, Wellesley College, Wellesley, MA 02181

1. Introduction. How do we best prepare mathematics students for careers in industry? How should we respond to students' requests for relevant courses? An informal survey suggests that solving these problems by decreasing the emphasis on classical pure mathematics is *not* in the students' long-term interest.

During a recent review of the undergraduate major at S.U.N.Y. at Stony Brook, 48 mathematicians in industry, business, and government replied to a questionnaire from the Department of Mathematics. The survey sought opinions on the relative merits, for students entering jobs with a bachelor's degree, of the traditional undergraduate curriculum, advanced topics in "pure" mathematics, computer programming, additional computer science, and specialized or applied topics. The questionnaire asked for comparisons between courses without limiting the total number that could be recommended.

The mathematicians who participated in the survey favored a broad background which combines a large dose of classical, pure mathematics with the problem-solving experience of applied courses and computer programming. The responses distinguished between various topics in both their potential value and whether they should be studied formally in courses. For example, a knowledge of computer programming through either course work or other experience was considered a basic skill, essential to the positions available today. Modeling, an important activity

Marcus, on the other hand, objects to the characterization of Blankinship's algorithm as "simpler," arguing that the two are significantly different, with the Blankinship algorithm having the advantage of applying to arbitrarily many integers rather than just two, of requiring no backtracking, and of relying on the familiar concept of row operations. His own algorithm, on the other hand, has the advantage of requiring that only the sequence of remainders be recorded and of producing the solutions by a single string of operations.

The publication of M. Hendy's note, "The retrial of the lower Slobbovian counterfeiters" (this MONTHLY, 87 (1980) 200–201), evoked for Juan Jorge Schäffer of Carnegie-Mellon University the memory that, when Problem E1096 appeared in this MONTHLY, 61 (1954) 46, he was a student in Zürich. P. Erdős, who happened to visit at that time, mentioned the problem to him, and they "soon found the fallacious 'solution' that had presumably been submitted in the meantime to the MONTHLY, and then discovered the fallacy, as well as the correct solution. He goes on to say that, when later that year he found that the fallacious "solution" had been published (this MONTHLY, 61 (1954) 472–473), he "wrote the editor of the Elementary Problems section to point out the error and provide a correct solution." When he received no acknowledgement, and the error was not corrected nor was a valid solution published, he did not follow up the matter. He encloses a copy of his 1954 letter "as a curiosity," having "nothing to complain about" and notes that "I shall be able to continue to boast that my Erdős number is 1, but that it is unlisted."

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

UNDERGRADUATE TRAINING FOR INDUSTRIAL CAREERS

ANN K. STEHNEY
Department of Mathematics, Wellesley College, Wellesley, MA 02181

1. Introduction. How do we best prepare mathematics students for careers in industry? How should we respond to students' requests for relevant courses? An informal survey suggests that solving these problems by decreasing the emphasis on classical pure mathematics is *not* in the students' long-term interest.

During a recent review of the undergraduate major at S.U.N.Y. at Stony Brook, 48 mathematicians in industry, business, and government replied to a questionnaire from the Department of Mathematics. The survey sought opinions on the relative merits, for students entering jobs with a bachelor's degree, of the traditional undergraduate curriculum, advanced topics in "pure" mathematics, computer programming, additional computer science, and specialized or applied topics. The questionnaire asked for comparisons between courses without limiting the total number that could be recommended.

The mathematicians who participated in the survey favored a broad background which combines a large dose of classical, pure mathematics with the problem-solving experience of applied courses and computer programming. The responses distinguished between various topics in both their potential value and whether they should be studied formally in courses. For example, a knowledge of computer programming through either course work or other experience was considered a basic skill, essential to the positions available today. Modeling, an important activity

in practice, was suggested for on-the-job learning. The theory of complex variables, less appropriate for independent study, was strongly recommended as coursework.

Support for programming and applied mathematics had been anticipated from the extensive work done by recent committees of the NRC [3] and CUPM [2]. The unexpected results of the Stony Brook survey were the preferred emphasis within these areas and the support for including the more theoretical courses. The respondents articulated a disturbing discrepancy between the courses which make a candidate attractive to an employer initially and the training most likely to correlate with success in the long run. These results contradict the usual assumptions about the needs of nonacademic employers, on whose behalf curriculum changes have been proposed and in whose name departments hope to attract students.

2. Computer Programming and Computer Science. Programming was described as an important tool, if not the main activity, of those hired today with a bachelor's degree. Two or three semesters of programming in a high-level language were recommended, with concepts such as structured programming included with instruction in syntax. Programming experience in mathematics courses, part-time or summer jobs, or internships would be an acceptable substitute for coursework. A systems programmer at a large computer company said that more than one programming language is "not a plus," since evidence of the ability to do the job (such as a good record in mathematics) would be evidence of the ability to learn other languages "on your own," as needed.

"Certainly knowledge of a programming language and some experience in programming is important," wrote a group supervisor in a computing research division. "My own requirements might stop there, but I believe most employers would be impressed by a considerable number of computer science courses." Well, some would and some would not. The question of computer science beyond programming—how much and what topics—elicited the most divergent opinions. A consulting company mathematician called such courses "of small value for most of our applications." At the other extreme, naturally enough, were the responses from employees of computer companies. One recommended a background in theoretical computer science and, in fact, would "prefer a straight C.S. degree." Another said there is a place for majors in pure mathematics who have an interest in computers, a programming course, and "a few 'advanced' computer science courses."

3. Applied Mathematics. Experience in "getting answers to specific problems" is the purpose of the two or three semesters of applied mathematics that were recommended. Except in fields requiring a working knowledge of statistics, the particular topic was less important to those surveyed than the practical experience of problem-solving. "Teach them to solve differential equations, matrix transformations, or whatever, but require them to get numerical answers that can be checked against actual experience," urged a senior advisor to a major aerospace company. Another consultant recommended "working on a real problem that requires the use of a computer and for which the straightforward textbook approach fails to give the right answer. Examples of such ideas include the inversion of a near singular matrix, solutions to ill-conditioned differential equations, or even the discrete time analysis of a continuous system."

Among applied courses, the preference was clearly for the more traditional topics—numerical analysis, statistics (using a major statistical package such as SPSS), probability, differential equations, numerical linear algebra, or optimization. Numerical analysis, strongly recommended for both computer science and applied mathematics, was seen as a rare opportunity to bring together these fields. There was widespread agreement that "knowledge in the area of operations research or modeling, although of interest, may be obtained once employed through either in-house training or short summer courses." A staff member at a research laboratory described operations research as a marketable skill, but one of the early practitioners in the field has "always thought it was better learned by apprenticeship; too much mathematical apparatus obscures the operational facts." A systems analyst at a national laboratory, who uses optimization

and simulation techniques, relies on measure theory, topology, and her undergraduate courses in calculus, elementary probability and statistics, symbolic logic, and programming.

4. Relations with Scientists and Engineers. Nonacademic work often involves collaboration with researchers in other fields. Courses in the physical sciences were recommended to develop the ability to communicate with and appreciate the problems of scientists and engineers. "Of course," wrote one research group supervisor, "many scientists know a great deal of mathematics, at least that of use to them. To be useful, a mathematician must exceed this knowledge in breadth or depth, or in both ways. This suggests to me that a mathematics major should take a great many mathematics courses, including those usually required of engineers, for instance, as well as the theoretical courses required in the normal mathematics curriculum."

5. Pure Mathematics. What of those theoretical courses? The questionnaire included several statements for the respondents to indicate a degree of agreement or disagreement. There was strong agreement with the statement "We prefer students with a broad background in classical mathematics [described as calculus, linear algebra, advanced calculus, one complex variable, etc.] and we will train them in more specialized knowledge." There was similarly strong disagreement with the statement "With respect to [our] positions, any mathematics beyond advanced calculus (e.g., one complex variable, group theory, topology, etc.) is not useful in an undergraduate education." Elaborating, the reaction to abstract algebra and topology ranged from "not particularly desirable" to "should be included," but mostly favorable. Symbolic logic was called "invaluable" to learn "ways to structure ideas at all levels, from abstract models to FORTRAN programs." A "sound foundation in set theory and linear algebra" was recommended by a senior computer scientist for all mathematics undergraduates.

Pure mathematics was generally credited with developing the ability to "think clearly," "abstractly," "analytically," and "precisely," and to "appreciate rigor." The issue was whether such experience is a "luxury," as one researcher called it, when the student needs the "ability to *do* something which industry is willing to pay for." The general response was that a thorough grounding in the classical material is (at least) as important as the immediately marketable skills developed in applied courses. "It is generally not possible to predict the sort of work one might be doing, even in a few years." Only a broad background including advanced work in pure mathematics was seen to provide the "flexibility to learn whatever else might become necessary" as technology and interests—one's own and one's employer's—inevitably change. There was little concern that the specific material of these courses would not actually be used; the experience of studying it and the added insight and mastery of more elementary concepts were adequate reason to include them.

The questionnaire did not explicitly address the issue of "continuous" versus "discrete" mathematics, but some of the replies did. Noting that a year of calculus is sufficient for their purposes, a data systems engineer and a systems analyst argued for replacing further study of continuous functions (advanced calculus, differential equations, analysis) by discrete mathematics (combinatorics, difference equations, graph theory, semi-groups, lattice theory). This is not, of course, a new idea (see [**1**] and [**4**], for example), but considering the rise of introductory "finite math" courses in the past two decades, it seemed to find little support among the respondents. For the most part, those who work with "structures" and discrete phenomena recommended the more traditional offerings in set theory, elementary probability and statistics, symbolic logic, and abstract algebra along with linear algebra and advanced calculus.

6. Conclusions. This survey sought representative opinions of nonacademic mathematicians as guidelines for revising the undergraduate mathematics program and advising students at S.U.N.Y. at Stony Brook. The recommended training combines experience in programming and problem-solving courses with a background in classical, pure mathematics. A good record in such a program is both marketable and flexible; neither aspect can be safely ignored. "The paradox,"

one respondent observed, "is that specialized courses (numerical analysis, optimization, numerical linear algebra) increase the probability of landing a job but a *broad* general background (topology, abstract algebra, complex analysis) is more important for long term success."

An informal study like this one cannot provide definitive answers to complex issues. It can, however, uncover opinions that have not received much attention otherwise. The results here, for example, challenge the assumptions that replacing "theory" by "practice" is sound preparation for a career. To do so might attract students to the department, but it is not clear that it will serve them well beyond a first job. In light of proposals for significant revisions in course offerings and major requirements, further evidence for or against these changes is clearly needed. If the opinions reported here are indeed representative of our colleagues in industry, the task ahead of the mathematical community is not to withdraw the theoretical part of the college curriculum but to educate career counselors, personnel officers, and students to the recommendations of those working in the field.

References

1. D. E. Knuth, Computer science and its relation to mathematics, this MONTHLY, 81 (1974) 323–343.
2. MAA Committee on the Undergraduate Program in Mathematics, Recommendations for a General Mathematical Sciences Program, Mathematical Association of America, Washington, D.C., 1981.
3. NRC Ad Hoc Committee on Applied Mathematics Training, The Role of Applications in the Undergraduate Mathematics Curriculum, National Academy of Sciences, Washington, D.C., 1979.
4. A. Ralston, Computer science, mathematics, and the undergraduate curricula in both, this MONTHLY, 88 (1981) 472–485.

MISCELLANEA

Our Favorite Things

110.

The algebraist:

Matrices, integers, rationals, reals;
Then polynomials and all their ideals;
Real-valued functions and similar things:
These are a few of my favorite rings.

The computer scientist:

Pascal and Fortran and languages formal;
Palindrome grammars in Backus-form-normal;
Post correspondence and regular things:
These are a few of my favorite strings.

(With apologies to Rodgers and Hammerstein)

—Robin J. Wilson
The Open University
Milton Keynes, England

—Ronald E. Prather
University of Denver
Denver, CO 80208

one respondent observed, "is that specialized courses (numerical analysis, optimization, numerical linear algebra) increase the probability of landing a job but a *broad* general background (topology, abstract algebra, complex analysis) is more important for long term success."

An informal study like this one cannot provide definitive answers to complex issues. It can, however, uncover opinions that have not received much attention otherwise. The results here, for example, challenge the assumptions that replacing "theory" by "practice" is sound preparation for a career. To do so might attract students to the department, but it is not clear that it will serve them well beyond a first job. In light of proposals for significant revisions in course offerings and major requirements, further evidence for or against these changes is clearly needed. If the opinions reported here are indeed representative of our colleagues in industry, the task ahead of the mathematical community is not to withdraw the theoretical part of the college curriculum but to educate career counselors, personnel officers, and students to the recommendations of those working in the field.

References

1. D. E. Knuth, Computer science and its relation to mathematics, this MONTHLY, 81 (1974) 323–343.

2. MAA Committee on the Undergraduate Program in Mathematics, Recommendations for a General Mathematical Sciences Program, Mathematical Association of America, Washington, D.C., 1981.

3. NRC Ad Hoc Committee on Applied Mathematics Training, The Role of Applications in the Undergraduate Mathematics Curriculum, National Academy of Sciences, Washington, D.C., 1979.

4. A. Ralston, Computer science, mathematics, and the undergraduate curricula in both, this MONTHLY, 88 (1981) 472–485.

MISCELLANEA

Our Favorite Things

110.

The algebraist:

Matrices, integers, rationals, reals;
Then polynomials and all their ideals;
Real-valued functions and similar things:
These are a few of my favorite rings.

The computer scientist:

Pascal and Fortran and languages formal;
Palindrome grammars in Backus-form-normal;
Post correspondence and regular things:
These are a few of my favorite strings.

(With apologies to Rodgers and Hammerstein)

—Robin J. Wilson
The Open University
Milton Keynes, England

—Ronald E. Prather
University of Denver
Denver, CO 80208

PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, DAVID BORWEIN (ADVANCED PROBLEMS), H. M. W. EDGAR (ELEMENTARY PROBLEMS), AND D. H. MUGLER

Send all ***proposed*** *problems, in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1984. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2974 [1982, 756] **Correction**. *Proposed by Jordi Dou, Barcelona, Spain.*

This problem, posed originally in the December 1982 issue of this MONTHLY, page 756, contained an error introduced in the editorial process. The problem should read: "Let AMB (oriented clockwise) and CMD (counter-clockwise) be similar triangles. Prove that triangles ACX (clockwise) and YDB (clockwise), both similar to the first triangles, have the same circumcenter."

E 3007. *Proposed by George Odom, Poughkeepsie, NY.*

Let A and B be the midpoints of the sides EF and ED of an equilateral triangle DEF. Extend AB to meet the circumcircle (of DEF) at C. Show that B divides AC according to the golden section.

E 3008. *Proposed by Roger Cuculière, Paris, France.*

Show that for $n \geqslant 2$ the polynomial $x^n - x^{n-1} - x^{n-2} - \cdots - x - 1$ is irreducible over the rationals.

E 3009. *Proposed by Chris Jantzen, University of Wisconsin.*

Points X, Y, Z are chosen on the sides of a triangle ABC such that

$$\frac{AX}{XB} = \frac{BY}{YC} = \frac{CZ}{ZA} = k$$

and a triangle PQR is formed using CX, AY and BZ as sides. The operation is repeated on the triangle PQR, that is, the points X', Y', Z' are chosen on the sides of PQR such that

$$\frac{PX'}{X'Q} = \frac{QY'}{Y'R} = \frac{RZ'}{Z'P} = k$$

and a triangle LMN is formed using RX', PY' and QZ' as sides. Show that LMN is similar to ABC and find the ratio of similarity.

E 3010. *Proposed by F. David Hammer, Santa Cruz, CA.*

Prove that for all $n \geqslant 0$ the relation

$$\left[n^{1/2} + (n+1)^{1/2} + (n+2)^{1/2}\right] = \left[(9n+8)^{1/2}\right]$$

holds. Here $[\cdot]$ denotes the greatest integer function.

E 3011. *Proposed by C. P. Popescu, Romania.*

Let X be a connected (path connected) topological space and let A, B be two nonvoid disjoined subsets of X such that boundary (A) = boundary (B). Is it true that $\bar{A} \cup B = X$?

E 3012. *Proposed by J. L. Selfridge, Mathematical Reviews.*

Show that $2^a - 1$ does not divide $3^a - 1$ if $a > 1$. More generally, $2^a - 1$ does not divide $3^b - 1$ when $a > 1$ and a and b have the same parity.

SOLUTIONS OF ELEMENTARY PROBLEMS

$an^2 + bn + c$ Is Always Divisible by a Prime $p \equiv 1 \bmod 4$

E 2883 [1981, 291]. *Proposed by J. M. Patin, St. Herblain, France.*

Can integers a, b, c be found so that the nonconstant polynomial $an^2 + bn + c$ has all its prime factors congruent to 3(mod 4) for $n = 1, 2, \ldots$?

Solution by Jacob A. Brandler, Brooklyn, NY; Robert Breusch, Amherst, Mass; Benny Cheng & Dinh Thế Hùng, students, University of California, Berkeley; Lorraine L. Foster, California State University, Northridge; Robert Gilmer, Florida State University; Sahib Singh, Clarion State College, PA; Allen Stenger, Hughes Aircraft Company; and University of South Alabama Problem Group. No such integers exist. Set $p(n) = an^2 + bn + c$, then $4ap(n) = (2an + b)^2 + 4ac - b^2$. If $a = 0$ or $4ac - b^2 = 0$, then $p(n)$ is divisible by primes of the form 1(mod 4) for infinitely many n so we assume that $a(4ac - b^2)$ is not zero and let $p_1, p_2, \ldots, p_m$ be the distinct odd prime divisors of $a(4ac - b^2)$. The Chinese Remainder Theorem and Dirichlet's Theorem guarantee that there is a prime $p > a$ such that $p \equiv 1(\bmod p_i)$ for all $i, 1 \leqslant i \leqslant m$, and $p \equiv 1(\bmod 8)$. Since $(p_i/p) = (p/p_i) = (2/p) = (-1/p) = 1$, there is an x such that $x^2 \equiv b^2 - 4ac(\bmod p)$. Since $(2a, p) = 1$, $2an + b \equiv x(\bmod p)$ has a solution for n, so $p | p(n)$ for some positive n.

Clearly there are infinitely many such primes p. Cheng and Hùng replaced 1 mod 4 by 1 mod m; Foster proved a considerable generalization. Advanced Problem 6340 (page 489) seems to be the last word in this connection.

Every prime divisor of $4n^2 + 1$ is congruent to 1 mod 4.

Also solved by the proposer.

Random Walks and Catalan Numbers

E 2903 [1981, 619]. *Proposed by Louis W. Shapiro, Howard University.*

Consider walks in the northeast quadrant which start at $(0,0)$ and such that each step is one unit east or north and the points $(1,1), (3,3), \ldots, (2k+1, 2k+1), \ldots$ are forbidden. How many paths are there to $(2n, 2n)$?

Many solvers used the method described below; some found the solution of an equivalent

and a triangle LMN is formed using RX', PY' and QZ' as sides. Show that LMN is similar to ABC and find the ratio of similarity.

E 3010. *Proposed by F. David Hammer, Santa Cruz, CA.*

Prove that for all $n \geqslant 0$ the relation

$$\left[n^{1/2} + (n+1)^{1/2} + (n+2)^{1/2}\right] = \left[(9n+8)^{1/2}\right]$$

holds. Here $[\cdot]$ denotes the greatest integer function.

E 3011. *Proposed by C. P. Popescu, Romania.*

Let X be a connected (path connected) topological space and let A, B be two nonvoid disjoined subsets of X such that boundary (A) = boundary (B). Is it true that $\bar{A} \cup B = X$?

E 3012. *Proposed by J. L. Selfridge, Mathematical Reviews.*

Show that $2^a - 1$ does not divide $3^a - 1$ if $a > 1$. More generally, $2^a - 1$ does not divide $3^b - 1$ when $a > 1$ and a and b have the same parity.

SOLUTIONS OF ELEMENTARY PROBLEMS

$an^2 + bn + c$ Is Always Divisible by a Prime $p \equiv 1 \bmod 4$

E 2883 [1981, 291]. *Proposed by J. M. Patin, St. Herblain, France.*

Can integers a, b, c be found so that the nonconstant polynomial $an^2 + bn + c$ has all its prime factors congruent to 3(mod 4) for $n = 1, 2, \ldots$?

Solution by Jacob A. Brandler, Brooklyn, NY; Robert Breusch, Amherst, Mass; Benny Cheng & Dinh Thế Hùng, students, University of California, Berkeley; Lorraine L. Foster, California State University, Northridge; Robert Gilmer, Florida State University; Sahib Singh, Clarion State College, PA; Allen Stenger, Hughes Aircraft Company; and University of South Alabama Problem Group. No such integers exist. Set $p(n) = an^2 + bn + c$, then $4ap(n) = (2an + b)^2 + 4ac - b^2$. If $a = 0$ or $4ac - b^2 = 0$, then $p(n)$ is divisible by primes of the form 1(mod 4) for infinitely many n so we assume that $a(4ac - b^2)$ is not zero and let $p_1, p_2, \ldots, p_m$ be the distinct odd prime divisors of $a(4ac - b^2)$. The Chinese Remainder Theorem and Dirichlet's Theorem guarantee that there is a prime $p > a$ such that $p \equiv 1 (\bmod\ p_i)$ for all $i, 1 \leqslant i \leqslant m$, and $p \equiv 1 (\bmod\ 8)$. Since $(p_i/p) = (p/p_i) = (2/p) = (-1/p) = 1$, there is an x such that $x^2 \equiv b^2 - 4ac (\bmod\ p)$. Since $(2a, p) = 1$, $2an + b \equiv x (\bmod\ p)$ has a solution for n, so $p \mid p(n)$ for some positive n.

Clearly there are infinitely many such primes p. Cheng and Hùng replaced 1 mod 4 by 1 mod m; Foster proved a considerable generalization. Advanced Problem 6340 (page 489) seems to be the last word in this connection.

Every prime divisor of $4n^2 + 1$ is congruent to 1 mod 4.

Also solved by the proposer.

Random Walks and Catalan Numbers

E 2903 [1981, 619]. *Proposed by Louis W. Shapiro, Howard University.*

Consider walks in the northeast quadrant which start at $(0,0)$ and such that each step is one unit east or north and the points $(1,1), (3,3), \ldots, (2k+1, 2k+1), \ldots$ are forbidden. How many paths are there to $(2n, 2n)$?

Many solvers used the method described below; some found the solution of an equivalent

and a triangle LMN is formed using RX', PY' and QZ' as sides. Show that LMN is similar to ABC and find the ratio of similarity.

E 3010. *Proposed by F. David Hammer, Santa Cruz, CA.*

Prove that for all $n \geqslant 0$ the relation

$$\left[n^{1/2} + (n+1)^{1/2} + (n+2)^{1/2}\right] = \left[(9n+8)^{1/2}\right]$$

holds. Here $[\cdot]$ denotes the greatest integer function.

E 3011. *Proposed by C. P. Popescu, Romania.*

Let X be a connected (path connected) topological space and let A, B be two nonvoid disjoined subsets of X such that boundary (A) = boundary (B). Is it true that $\bar{A} \cup B = X$?

E 3012. *Proposed by J. L. Selfridge, Mathematical Reviews.*

Show that $2^a - 1$ does not divide $3^a - 1$ if $a > 1$. More generally, $2^a - 1$ does not divide $3^b - 1$ when $a > 1$ and a and b have the same parity.

SOLUTIONS OF ELEMENTARY PROBLEMS

$an^2 + bn + c$ Is Always Divisible by a Prime $p \equiv 1 \bmod 4$

E 2883 [1981, 291]. *Proposed by J. M. Patin, St. Herblain, France.*

Can integers a, b, c be found so that the nonconstant polynomial $an^2 + bn + c$ has all its prime factors congruent to 3(mod 4) for $n = 1, 2, \ldots$?

Solution by Jacob A. Brandler, Brooklyn, NY; Robert Breusch, Amherst, Mass; Benny Cheng & Dinh Thế Hùng, students, University of California, Berkeley; Lorraine L. Foster, California State University, Northridge; Robert Gilmer, Florida State University; Sahib Singh, Clarion State College, PA; Allen Stenger, Hughes Aircraft Company; and University of South Alabama Problem Group. No such integers exist. Set $p(n) = an^2 + bn + c$, then $4ap(n) = (2an + b)^2 + 4ac - b^2$. If $a = 0$ or $4ac - b^2 = 0$, then $p(n)$ is divisible by primes of the form 1(mod 4) for infinitely many n so we assume that $a(4ac - b^2)$ is not zero and let $p_1, p_2, \ldots, p_m$ be the distinct odd prime divisors of $a(4ac - b^2)$. The Chinese Remainder Theorem and Dirichlet's Theorem guarantee that there is a prime $p > a$ such that $p \equiv 1(\bmod p_i)$ for all $i, 1 \leqslant i \leqslant m$, and $p \equiv 1(\bmod 8)$. Since $(p_i/p) = (p/p_i) = (2/p) = (-1/p) = 1$, there is an x such that $x^2 \equiv b^2 - 4ac(\bmod p)$. Since $(2a, p) = 1$, $2an + b \equiv x(\bmod p)$ has a solution for n, so $p | p(n)$ for some positive n.

Clearly there are infinitely many such primes p. Cheng and Hùng replaced 1 mod 4 by 1 mod m; Foster proved a considerable generalization. Advanced Problem 6340 (page 489) seems to be the last word in this connection.

Every prime divisor of $4n^2 + 1$ is congruent to 1 mod 4.

Also solved by the proposer.

Random Walks and Catalan Numbers

E 2903 [1981, 619]. *Proposed by Louis W. Shapiro, Howard University.*

Consider walks in the northeast quadrant which start at $(0,0)$ and such that each step is one unit east or north and the points $(1,1), (3,3), \ldots, (2k+1, 2k+1), \ldots$ are forbidden. How many paths are there to $(2n, 2n)$?

Many solvers used the method described below; some found the solution of an equivalent

problem in W. Feller's *Introduction to Probability*. The Catalan numbers give the number of ways of parenthesizing a product.

Solution by Jo Anne Simpson Growney, Bloomsburg State College, Pennsylvania. After some brief historical remarks, we shall show that the number of such paths, for $n = 0, 1, 2, \ldots$, is

$$C(2n) = \frac{1}{2n+1}\binom{4n}{2n}$$

where the symbol $C(2n)$ denotes the $2n$th member of the Catalan sequence.

In an article entitled "Arrangements of m things of one sort and n things of another sort under certain conditions of priority" (*Messenger of Math.*, vol. 8 (1878), pp. 105–114), W. A. Whitworth showed that $\frac{m-n+1}{m+1}\binom{m+n}{n}$ gives the number of routes from the origin (0,0) to the point (m, n) in the Cartesian plane which do not go above the line $y = x$ and which consist of m eastward paces and n northern paces ($m \geqslant n$) each one unit in length. In the special case $m = n$, Whitworth's formula gives the nth Catalan number.

The Catalan numbers have a long history dating back to Euler (1758). Their frequent occurrence in the mathematical literature has been well documented in a *Research Bibliography* compiled by H. W. Gould (West Virginia University, Morgantown, WV 25606).

To show that the problem given above is equivalent to Whitworth's problem, we need to make two simple observations:

(1) The restriction, in the given problem, that points with odd and equal coordinates be avoided means that, as a walk from $(0,0)$ to $(2n, 2n)$ progresses, at the $2n$ points $(0,1), (1,0), (2,3), (3,2), \ldots, (2n-2, 2n-1), (2n-1, 2n-2)$ there is only one possible choice for the next step.

(2) The restriction, in Whitworth's problem, that paths from $(0,0)$ to $(2n, 2n)$ not go above the diagonal, means that at the $2n$ points $(0,0), (1,1), (2,2), \ldots, (2n-1, 2n-1)$ there is only one possible choice for the next step.

Thus the given problem and Whitworth's problem both ask us to find the number of different sequences of $2n$ paces—n of which are eastward and n of which are northward—which have an identical number of options excluded. The solutions are therefore equivalent.

A proof that the nth Catalan number counts the number of paths from $(0,0)$ to (n, n) that do not cross the diagonal may be obtained by (a quite laborious) mathematical induction—based on the demonstration that the number of paths to (n, n) is $\frac{4n-2}{n+1}$ times the number of such paths to $(n-1, n-1)$.

Alternatively, use of the recursion relation

$$C(n) = C(0)C(n-1) + C(1)C(n-2) + \cdots + C(n-1)C(0), \text{with } C(0) = 0,$$

(which can readily be seen to be true for northeast quadrant paths of Whitworth's type) provides another basis for derivation of the formula

$$C(n) = \frac{1}{n+1}\binom{2n}{n}.$$

Also solved by D. M. Bloom, P. S. Bruckman, K. M. Chan and Y. M. Chow (Singapore), F. E. Crawford (student), J. Dou (Spain), L. L. Foster, I. Gessel, R. Z. Goldstein, J. P. Goulden (Canada), V. Hernandez (Spain), D. R. Hill and R. W. Irving (UK), L.-H. Hsu (Taiwan), O. Jorsboe (Denmark), O. Krafft (Germany), O. P. Lossers (Netherlands), W. A. Newcomb, D. E. Orr, C. R. Pranesachar (India), R. P. Savage, Jr., R. Scharlach (Germany), R. Sheets, A. Smuckler (Israel), D. G. Weinman, M. Woltermann, K. Zikan, and the proposer.

Representation of lim A^n (of a Matrix)

E 2907 [1981, 618]. *Proposed by H. Kestelman, University College, London.*

Given that the sequence $A, A^2, A^3, \ldots$ converges to a nonzero matrix A^∞, show that $A^\infty = V(WV)^{-1}W$ where V is any matrix whose columns are a basis of the right kernel of $A - I$, and W is any one whose rows are a basis of the left kernel of $A - I$.

Solution by Emeric Deutsch, Polytechnic Institute of New York. It is known (see U. G. Rothblum, *Mathematical Programming Study*, vol. 6, 1976, pp. 188–201), and relatively easy to verify, that if E is an idempotent matrix and the columns of a matrix X (resp. Y) form a basis of range E (resp. range E^*), then

$$E = X(Y^*X)^{-1}Y^*. \tag{1}$$

By assumption, we have $AV = V$, $WA = W$, which imply at once that $A^\infty V = V$ and $WA^\infty = W$. It is also easy to see that A^∞ is idempotent. Thus, the columns of V (resp. W^*) form a basis of range A^∞ (resp. range $(A^\infty)^*$). Then, from (1) we obtain the desired equality.

Also solved by E. Deutsch (second solution), A. L. Duarte and J. F. Quierò (Portugal), D. Ẹ. Orr, P.-Y. Wu (Taiwan), and the proposer.

A Polynomial Congruence Modulo Higher and Higher Powers

E 2908 [1981, 705]. *Proposed by Ken Brown, Nova High School, Fort Lauderdale, FL.*

Let $p(x) = 2x^3 + 3x^2 - 1$. Define the sequence $\{a_n\}$ by

$$a_1 = 4, a_{n+1} = p(a_n) - 10^f\left[p(a_n)/10^f\right], \quad n > 0, \quad f = 2^n.$$

Show that a_n satisfies the congruence $x^3 \equiv x \pmod{10^e}$, $e = 2^{n-1}$.

Solution by Ira Gessel, Massachusetts Institute of Technology, and University of South Alabama Problem Group. It can be shown by induction that mod 5^e, $a_n \equiv -1$ if n is odd; $a_n \equiv 0$ if n is even. Also mod 2^{2e}, $a_n \equiv 0$ if n is odd; $a_n \equiv -1$ if n is even. Thus if m is odd, $a_n^m \equiv a_n \pmod{10^e}$.

O. P. Lossers and C-N. Lee showed that $a_n^2 + a_n \equiv 0 \pmod{10^e}$. L. L. Foster gave further generalizations.

Also solved by an anonymous Chinese solver, L. L. Foster, P. Kumar (India), C-N. Lee, O. P. Lossers (Netherlands), A. Stenger, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1984. The solver's full post-office address should be on each sheet.

6436. *Proposed by H. Kestelman, University College, London.*

Let V be a real vector space normed by a scalar product. Let $f\colon V \to V$ be such that whenever a, b, c are vertices of a triangle right-angled at a, then $f(a), f(b), f(c)$ are vertices of a triangle right-angled at $f(a)$. Show that f is a continuous collineation.

6437. *Proposed by Bhaskar Bagchi, Indian Statistical Institute, Calcutta.*

Given n meromorphic functions $f_1, f_2, \ldots, f_n$ on a planar region U, show that there exists a holomorphic function g on U such that $f_j(z) \neq g(z)$ for $1 \leqslant j \leqslant n$ and for all z in U.

Representation of lim A^n (of a Matrix)

E 2907 [1981, 618]. *Proposed by H. Kestelman, University College, London.*

Given that the sequence $A, A^2, A^3, \ldots$ converges to a nonzero matrix A^∞, show that $A^\infty = V(WV)^{-1}W$ where V is any matrix whose columns are a basis of the right kernel of $A - I$, and W is any one whose rows are a basis of the left kernel of $A - I$.

Solution by Emeric Deutsch, Polytechnic Institute of New York. It is known (see U. G. Rothblum, *Mathematical Programming Study*, vol. 6, 1976, pp. 188–201), and relatively easy to verify, that if E is an idempotent matrix and the columns of a matrix X (resp. Y) form a basis of range E (resp. range E^*), then

$$E = X(Y^*X)^{-1}Y^*. \tag{1}$$

By assumption, we have $AV = V$, $WA = W$, which imply at once that $A^\infty V = V$ and $WA^\infty = W$. It is also easy to see that A^∞ is idempotent. Thus, the columns of V (resp. W^*) form a basis of range A^∞ (resp. range $(A^\infty)^*$). Then, from (1) we obtain the desired equality.

Also solved by E. Deutsch (second solution), A. L. Duarte and J. F. Quierò (Portugal), D. Ẹ. Orr, P.-Y. Wu (Taiwan), and the proposer.

A Polynomial Congruence Modulo Higher and Higher Powers

E 2908 [1981, 705]. *Proposed by Ken Brown, Nova High School, Fort Lauderdale, FL.*

Let $p(x) = 2x^3 + 3x^2 - 1$. Define the sequence $\{a_n\}$ by

$$a_1 = 4, a_{n+1} = p(a_n) - 10^f[p(a_n)/10^f], \quad n > 0, \quad f = 2^n.$$

Show that a_n satisfies the congruence $x^3 \equiv x \pmod{10^e}$, $e = 2^{n-1}$.

Solution by Ira Gessel, Massachusetts Institute of Technology, and University of South Alabama Problem Group. It can be shown by induction that mod 5^e, $a_n \equiv -1$ if n is odd; $a_n \equiv 0$ if n is even. Also mod 2^{2e}, $a_n \equiv 0$ if n is odd; $a_n \equiv -1$ if n is even. Thus if m is odd, $a_n^m \equiv a_n \pmod{10^e}$.

O. P. Lossers and C-N. Lee showed that $a_n^2 + a_n \equiv 0 \pmod{10^e}$. L. L. Foster gave further generalizations.

Also solved by an anonymous Chinese solver, L. L. Foster, P. Kumar (India), C-N. Lee, O. P. Lossers (Netherlands), A. Stenger, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1984. The solver's full post-office address should be on each sheet.

6436. *Proposed by H. Kestelman, University College, London.*

Let V be a real vector space normed by a scalar product. Let $f\colon V \to V$ be such that whenever a, b, c are vertices of a triangle right-angled at a, then $f(a), f(b), f(c)$ are vertices of a triangle right-angled at $f(a)$. Show that f is a continuous collineation.

6437. *Proposed by Bhaskar Bagchi, Indian Statistical Institute, Calcutta.*

Given n meromorphic functions $f_1, f_2, \ldots, f_n$ on a planar region U, show that there exists a holomorphic function g on U such that $f_j(z) \not\equiv g(z)$ for $1 \leqslant j \leqslant n$ and for all z in U.

Representation of lim A^n (of a Matrix)

E 2907 [1981, 618]. *Proposed by H. Kestelman, University College, London.*

Given that the sequence $A, A^2, A^3, \ldots$ converges to a nonzero matrix A^∞, show that $A^\infty = V(WV)^{-1}W$ where V is any matrix whose columns are a basis of the right kernel of $A - I$, and W is any one whose rows are a basis of the left kernel of $A - I$.

Solution by Emeric Deutsch, Polytechnic Institute of New York. It is known (see U. G. Rothblum, *Mathematical Programming Study*, vol. 6, 1976, pp. 188–201), and relatively easy to verify, that if E is an idempotent matrix and the columns of a matrix X (resp. Y) form a basis of range E (resp. range E^*), then

$$E = X(Y^*X)^{-1}Y^*. \tag{1}$$

By assumption, we have $AV = V$, $WA = W$, which imply at once that $A^\infty V = V$ and $WA^\infty = W$. It is also easy to see that A^∞ is idempotent. Thus, the columns of V (resp. W^*) form a basis of range A^∞ (resp. range $(A^\infty)^*$). Then, from (1) we obtain the desired equality.

Also solved by E. Deutsch (second solution), A. L. Duarte and J. F. Quieró (Portugal), D. Ẹ. Orr, P.-Y. Wu (Taiwan), and the proposer.

A Polynomial Congruence Modulo Higher and Higher Powers

E 2908 [1981, 705]. *Proposed by Ken Brown, Nova High School, Fort Lauderdale, FL.*

Let $p(x) = 2x^3 + 3x^2 - 1$. Define the sequence $\{a_n\}$ by

$$a_1 = 4, a_{n+1} = p(a_n) - 10^f[p(a_n)/10^f], \quad n > 0, \quad f = 2^n.$$

Show that a_n satisfies the congruence $x^3 \equiv x \pmod{10^e}$, $e = 2^{n-1}$.

Solution by Ira Gessel, Massachusetts Institute of Technology, and University of South Alabama Problem Group. It can be shown by induction that mod 5^e, $a_n \equiv -1$ if n is odd; $a_n \equiv 0$ if n is even. Also mod 2^{2e}, $a_n \equiv 0$ if n is odd; $a_n \equiv -1$ if n is even. Thus if m is odd, $a_n^m \equiv a_n \pmod{10^e}$.

O. P. Lossers and C-N. Lee showed that $a_n^2 + a_n \equiv 0 \pmod{10^e}$. L. L. Foster gave further generalizations.

Also solved by an anonymous Chinese solver, L. L. Foster, P. Kumar (India), C-N. Lee, O. P. Lossers (Netherlands), A. Stenger, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1984. The solver's full post-office address should be on each sheet.

6436. *Proposed by H. Kestelman, University College, London.*

Let V be a real vector space normed by a scalar product. Let $f\colon V \to V$ be such that whenever a, b, c are vertices of a triangle right-angled at a, then $f(a), f(b), f(c)$ are vertices of a triangle right-angled at $f(a)$. Show that f is a continuous collineation.

6437. *Proposed by Bhaskar Bagchi, Indian Statistical Institute, Calcutta.*

Given n meromorphic functions $f_1, f_2, \ldots, f_n$ on a planar region U, show that there exists a holomorphic function g on U such that $f_j(z) \neq g(z)$ for $1 \leqslant j \leqslant n$ and for all z in U.

6438. *Proposed by J. L. Brenner, Palo Alto, California.*

Construct a differentiable real function that is positive at every rational point, but zero at an uncountable number of points, or prove that no such function exists.

SOLUTIONS OF ADVANCED PROBLEMS

Three Circles with Collinear Centers

3887 [1938, 482]. *Proposed by V. Thébault, La Mans, France*

Through the vertex A of a triangle ABC, a straight line AM is drawn cutting the side BC in M. Let 2θ be the angle AMC; O and I the centers of the circumscribed circle (O) and the inscribed circle (I) of ABC. The circles (ω_1) and (ω_2) with centers ω_1 and ω_2 and radii ρ_1 and ρ_2 are each tangent to (O) and the first is tangent also to the two sides of angle AMC while the second is tangent to the two sides of angle AMB. Prove that: (1) The straight line joining ω_1 and ω_2 passes through I. (2) The point I divides the segment $\omega_1\omega_2$ in the ratio $\tan^2\theta : 1$; and $\rho_1 + \rho_2 = r^2 \sec^2\theta$, where r is the radius of (I). (See Fig. 1.)

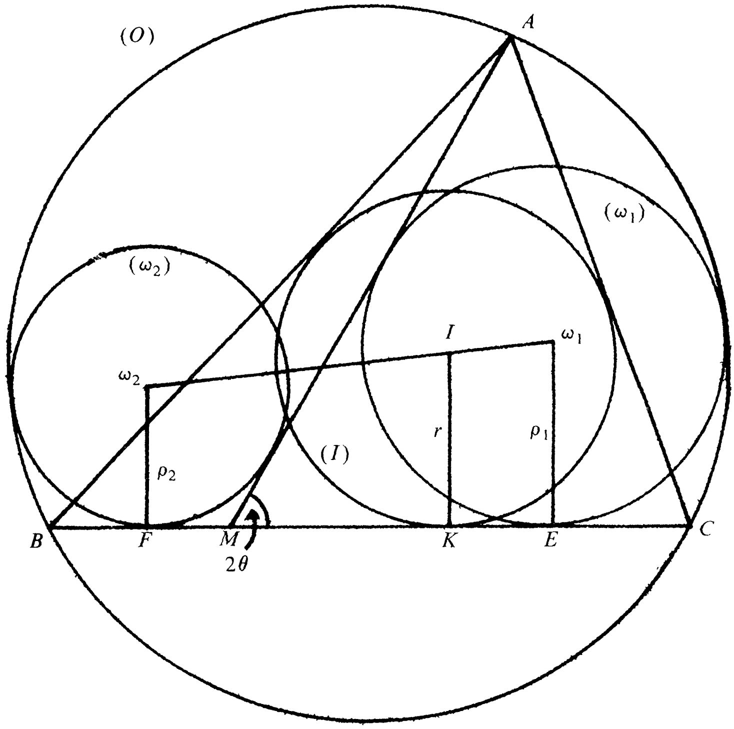


FIG. 1.

Editorial comment. The above problem, proposed over 40 years ago by the prolific and penetrating geometer V. Thébault, has remained unsolved until recently, when K. B. Taylor of England submitted a solution. Mr. Taylor had run across a restatement of the problem in C. Stanley Ogilvy: *Tomorrow's Math, Unsolved Problems for the Amateur,* Oxford University Press, 1962, p. 70. The solution given by Mr. Taylor occupies 24 manuscript pages and is based upon a lengthy calculation of involved distances. For lack of space, and with apologies to Mr. Taylor, we here reproduce his solution in a brief summarized form. Thébault's finding can be more

6438. *Proposed by J. L. Brenner, Palo Alto, California.*

Construct a differentiable real function that is positive at every rational point, but zero at an uncountable number of points, or prove that no such function exists.

SOLUTIONS OF ADVANCED PROBLEMS

Three Circles with Collinear Centers

3887 [1938, 482]. *Proposed by V. Thébault, La Mans, France*

Through the vertex A of a triangle ABC, a straight line AM is drawn cutting the side BC in M. Let 2θ be the angle AMC; O and I the centers of the circumscribed circle (O) and the inscribed circle (I) of ABC. The circles (ω_1) and (ω_2) with centers ω_1 and ω_2 and radii ρ_1 and ρ_2 are each tangent to (O) and the first is tangent also to the two sides of angle AMC while the second is tangent to the two sides of angle AMB. Prove that: (1) The straight line joining ω_1 and ω_2 passes through I. (2) The point I divides the segment $\omega_1\omega_2$ in the ratio $\tan^2\theta : 1$; and $\rho_1 + \rho_2 = r^2 \sec^2\theta$, where r is the radius of (I). (See Fig. 1.)

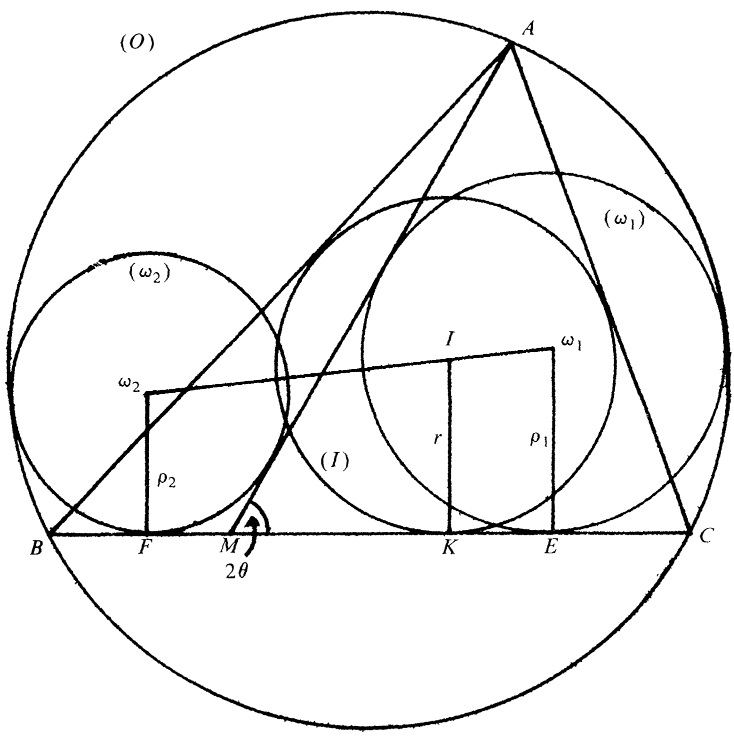


FIG. 1.

Editorial comment. The above problem, proposed over 40 years ago by the prolific and penetrating geometer V. Thébault, has remained unsolved until recently, when K. B. Taylor of England submitted a solution. Mr. Taylor had run across a restatement of the problem in C. Stanley Ogilvy: *Tomorrow's Math, Unsolved Problems for the Amateur*, Oxford University Press, 1962, p. 70. The solution given by Mr. Taylor occupies 24 manuscript pages and is based upon a lengthy calculation of involved distances. For lack of space, and with apologies to Mr. Taylor, we here reproduce his solution in a brief summarized form. Thébault's finding can be more

attractively stated as follows: Let P be a quadrilateral inscribed in a circle (O) and let Q be the quadrilateral formed by the centers of the four circles internally touching (O) and each of the two diagonals of P. Then the incenters of the four triangles having for sides the sides and diagonals of P form a rectangle R inscribed in Q.

Solution by K. B. Taylor, Salford, England. Let E, F, K be the points of tangency of $(\omega_1), (\omega_2), (I)$ with line BC. By brute calculation we find, using standard notation applied to triangle ABC,

$$\rho_1 = r\{-r\tan^2\theta - (c-b)\tan\theta + r_1\}/(r_1 - r),$$
$$\rho_2 = r\{-r\cot^2\theta + (c-b)\cot\theta + r_1\}/(r_1 - r),$$
$$FK = r\cot\theta, \quad KE = r\tan\theta.$$

Using these expressions, one can show that

$$\text{slope } \omega_2 I = \text{slope } I\omega_1,$$

whence ω_1, I, ω_2 are collinear. One can also show that

$$\omega_2 I : I\omega_1 = FK : KE = 1 : \tan^2\theta,$$

and that

$$r = \rho_1\cos^2\theta + \rho_2\sin^2\theta.$$

(The relation $\rho_1 + \rho_2 = r^2\sec^2\theta$ given by the proposer is manifestly false, since it is dimensionally incorrect.)

From expressions similar to the above distance relations, one can show that if I' is the excenter opposite vertex A, (ω_1') and (ω_2') are the circles externally tangent to (O), both also tangent to BC and AM produced, then ω_1', I', ω_2' are collinear, and

$$\omega_1' I' : I'\omega_2' = \tan^2\theta : 1, \; r_1 = \rho_1'\cos^2\theta + \rho_2'\sin^2\theta.$$

When M is coincident with the point K of tangency of the incircle (I) with BC, then $\rho_1' = \rho_2' = r_1$ and $\tan 2\theta = 2r_1/(c-b)$. When M is coincident with K', $\rho_1 = \rho_2 = r$ and $\tan 2\theta = 2r/(c-b)$.

Borel Subsets of a Product Space

6023* [1975, 308; 1983, 136]. *Proposed by S. J. Sidney, University of Connecticut.*

If for each k in the uncountable index set K, I_k denotes a copy of $[0,1]$ and U_k denotes the copy of $(0,1]$ contained therein, prove or disprove that $\Pi_k U_k$ is a Borel set in the compact space $\Pi_k I_k$.

A solution by M. G. Pelling was published in the February 1983 issue of this MONTHLY. A. H. Stone (The University of Rochester) has pointed out that in a paper by D. Maharam and himself, Borel Boxes, Pacific J. Math., 81 (1979) 471–473, it is shown that a nonempty product $\Pi\{Y_k: k \in K\}$ is Borel in $\Pi\{X_k: k \in K\}$, where each X_k is first countable and a Hausdorff space, if and only if (1) each Y_k is Borel in X_k, and (2) for all but countably many k's, Y_k is closed in X_k.

This immediately yields the negative answer to the problem and the proof is simpler than Pelling's.

Additive Set Functions of Bounded Variation

6256 [1979, 132]. *Proposed by A. Kussmaul and P. E. Kopp, University of Hull, England.*

Prove or disprove the assertion in a recent text that every countably additive real-valued set

attractively stated as follows: Let P be a quadrilateral inscribed in a circle (O) and let Q be the quadrilateral formed by the centers of the four circles internally touching (O) and each of the two diagonals of P. Then the incenters of the four triangles having for sides the sides and diagonals of P form a rectangle R inscribed in Q.

Solution by K. B. Taylor, Salford, England. Let E, F, K be the points of tangency of $(\omega_1), (\omega_2), (I)$ with line BC. By brute calculation we find, using standard notation applied to triangle ABC,

$$\rho_1 = r\{-r\tan^2\theta - (c-b)\tan\theta + r_1\}/(r_1 - r),$$
$$\rho_2 = r\{-r\cot^2\theta + (c-b)\cot\theta + r_1\}/(r_1 - r),$$
$$FK = r\cot\theta, \quad KE = r\tan\theta.$$

Using these expressions, one can show that

$$\text{slope } \omega_2 I = \text{slope } I\omega_1,$$

whence ω_1, I, ω_2 are collinear. One can also show that

$$\omega_2 I : I\omega_1 = FK : KE = 1 : \tan^2\theta,$$

and that

$$r = \rho_1\cos^2\theta + \rho_2\sin^2\theta.$$

(The relation $\rho_1 + \rho_2 = r^2\sec^2\theta$ given by the proposer is manifestly false, since it is dimensionally incorrect.)

From expressions similar to the above distance relations, one can show that if I' is the excenter opposite vertex A, (ω_1') and (ω_2') are the circles externally tangent to (O), both also tangent to BC and AM produced, then ω_1', I', ω_2' are collinear, and

$$\omega_1' I' : I'\omega_2' = \tan^2\theta : 1, \; r_1 = \rho_1'\cos^2\theta + \rho_2'\sin^2\theta.$$

When M is coincident with the point K of tangency of the incircle (I) with BC, then $\rho_1' = \rho_2' = r_1$ and $\tan 2\theta = 2r_1/(c-b)$. When M is coincident with K', $\rho_1 = \rho_2 = r$ and $\tan 2\theta = 2r/(c-b)$.

Borel Subsets of a Product Space

6023* [1975, 308; 1983, 136]. *Proposed by S. J. Sidney, University of Connecticut.*

If for each k in the uncountable index set K, I_k denotes a copy of $[0,1]$ and U_k denotes the copy of $(0,1]$ contained therein, prove or disprove that $\Pi_k U_k$ is a Borel set in the compact space $\Pi_k I_k$.

A solution by M. G. Pelling was published in the February 1983 issue of this MONTHLY. A. H. Stone (The University of Rochester) has pointed out that in a paper by D. Maharam and himself, Borel Boxes, Pacific J. Math., 81 (1979) 471–473, it is shown that a nonempty product $\Pi\{Y_k: k \in K\}$ is Borel in $\Pi\{X_k: k \in K\}$, where each X_k is first countable and a Hausdorff space, if and only if (1) each Y_k is Borel in X_k, and (2) for all but countably many k's, Y_k is closed in X_k.

This immediately yields the negative answer to the problem and the proof is simpler than Pelling's.

Additive Set Functions of Bounded Variation

6256 [1979, 132]. *Proposed by A. Kussmaul and P. E. Kopp, University of Hull, England.*

Prove or disprove the assertion in a recent text that every countably additive real-valued set

attractively stated as follows: Let P be a quadrilateral inscribed in a circle (O) and let Q be the quadrilateral formed by the centers of the four circles internally touching (O) and each of the two diagonals of P. Then the incenters of the four triangles having for sides the sides and diagonals of P form a rectangle R inscribed in Q.

Solution by K. B. Taylor, Salford, England. Let E, F, K be the points of tangency of $(\omega_1), (\omega_2), (I)$ with line BC. By brute calculation we find, using standard notation applied to triangle ABC,

$$\rho_1 = r\{-r\tan^2\theta - (c-b)\tan\theta + r_1\}/(r_1 - r),$$
$$\rho_2 = r\{-r\cot^2\theta + (c-b)\cot\theta + r_1\}/(r_1 - r),$$
$$FK = r\cot\theta, \quad KE = r\tan\theta.$$

Using these expressions, one can show that

$$\text{slope } \omega_2 I = \text{slope } I\omega_1,$$

whence ω_1, I, ω_2 are collinear. One can also show that

$$\omega_2 I : I\omega_1 = FK : KE = 1 : \tan^2\theta,$$

and that

$$r = \rho_1\cos^2\theta + \rho_2\sin^2\theta.$$

(The relation $\rho_1 + \rho_2 = r^2\sec^2\theta$ given by the proposer is manifestly false, since it is dimensionally incorrect.)

From expressions similar to the above distance relations, one can show that if I' is the excenter opposite vertex A, (ω_1') and (ω_2') are the circles externally tangent to (O), both also tangent to BC and AM produced, then ω_1', I', ω_2' are collinear, and

$$\omega_1' I' : I'\omega_2' = \tan^2\theta : 1, \; r_1 = \rho_1'\cos^2\theta + \rho_2'\sin^2\theta.$$

When M is coincident with the point K of tangency of the incircle (I) with BC, then $\rho_1' = \rho_2' = r_1$ and $\tan 2\theta = 2r_1/(c-b)$. When M is coincident with K', $\rho_1 = \rho_2 = r$ and $\tan 2\theta = 2r/(c-b)$.

Borel Subsets of a Product Space

6023* [1975, 308; 1983, 136]. *Proposed by S. J. Sidney, University of Connecticut.*

If for each k in the uncountable index set K, I_k denotes a copy of $[0,1]$ and U_k denotes the copy of $(0,1]$ contained therein, prove or disprove that $\Pi_k U_k$ is a Borel set in the compact space $\Pi_k I_k$.

A solution by M. G. Pelling was published in the February 1983 issue of this MONTHLY. A. H. Stone (The University of Rochester) has pointed out that in a paper by D. Maharam and himself, Borel Boxes, Pacific J. Math., 81 (1979) 471–473, it is shown that a nonempty product $\Pi\{Y_k: k \in K\}$ is Borel in $\Pi\{X_k: k \in K\}$, where each X_k is first countable and a Hausdorff space, if and only if (1) each Y_k is Borel in X_k, and (2) for all but countably many k's, Y_k is closed in X_k.

This immediately yields the negative answer to the problem and the proof is simpler than Pelling's.

Additive Set Functions of Bounded Variation

6256 [1979, 132]. *Proposed by A. Kussmaul and P. E. Kopp, University of Hull, England.*

Prove or disprove the assertion in a recent text that every countably additive real-valued set

function on a ring R of sets is of bounded variation. Is the assertion true if R is an algebra of sets?

Solved by A. Brunnschweiler (Switzerland), F. S. Cater, A. Janssen (West Germany), R. Bruce Mericle, Charles Riley, J. C. Smit (Netherlands) and the proposers. All solvers gave an example where R is the ring of subsets A of a set X such that either A or $X - A$ is finite, defining $\mu(A)$ to be the cardinal of A for A finite, and minus the cardinal of $X - A$ for $X - A$ finite.

A Condition on Entire Functions

6279 [1979, 793; 1981, 542]. *Proposed by Lee A. Rubel, University of Illinois, Urbana-Champaign.*

Let $f(z)$ be an entire function such that the maximum modulus over every closed line segment L is achieved at one of the endpoints a and b of L; that is,

$$\max\{|f(z)|: z \in L\} = \max\{|f(a)|, |f(b)|\}.$$

Prove that $f(z)$ has either the form $A(z - B)^n$ or the form $A \exp Bz$, where A and B are constants and n is a nonnegative integer.

The proposer pointed out that there is a gap in the published solution. Below are two correct solutions.

I. *Solution by P. R. Chernoff and J. Essick, University of California, Berkeley.* If f has more than one zero, then $|f|$ vanishes on the segment joining two of its zeros; so f is identically zero. Assume henceforth that f is not identically zero.

Case 1. f has one zero. Without loss of generality we may suppose $f(0) = 0$. If f is a polynomial, it is necessarily of the form cz^n. But if f is not a polynomial, then by the Casorati-Weierstrass theorem there is a sequence of points p_n with $|p_n| \to \infty$ and $|f(p_n)| = \varepsilon_n \to 0$. By passing to a subsequence, we may suppose that the unit complex numbers $\omega_n = p_n/|p_n|$ converge to a limit ω. Then $|f(z)| \leqslant \varepsilon_n$ on the line segment joining 0 to p_n, so that $|f(z)|$ must be identically 0 on the ray through 0 along direction ω. This is a contradiction.

Case 2. f has no zeros. Then $f = e^g$ for an entire function g. The real part $u(z)$ of $g(z)$, when restricted to any line L, can have no local maxima. Consequently, if the directional derivative $D_\omega u(a)$ along the direction ω at the point a should be positive, it follows that $u(a + t\omega)$ must be nondecreasing for $t \geqslant 0$, and thus $u(z) \geqslant u(a)$ for all z on the ray emanating from a along direction ω. Note that $D_\omega u(a) = \mathrm{Re}(\omega g'(a))$.

We claim that g' is constant. If not, we can obviously find three points a_1, a_2, a_3 such that the numbers $c_k = g'(a_k)$ are very close to ω_0^k, $k = 1, 2, 3$, where $\omega_0 = e^{2\pi i/3}$. Suppose now that ω is any complex number with $\mathrm{Re}(\omega c_k) > 0$. Then $D_\omega u(a_k) > 0$, so that $u(z) \geqslant u(a_k)$ for all z in the half plane

$$P_k = \{z : \mathrm{Re}((z - a_k)c_k)\} \geqslant 0.$$

Now, by the choice of c_1, c_2, c_3, the union $P_1 \cup P_2 \cup P_3$ of these half planes includes all of $\mathbb{C}$ except possibly a bounded triangular region, and therefore $u(z)$ is bounded below on the entire plane. This means that $g(z)$ must be constant, contradicting our assumption about g'. Thus g' is a constant A, $g(z) = Az + B$, and $f(z) = \exp(Az + B)$.

II. *Solution by Jan Maly and Zdenek Vlasek, Charles University, Prague.* If f vanishes at the endpoints of some segment $L = [a, b]$, then $f \equiv 0$ on L, and therefore $f \equiv 0$ on $\mathbb{C}$. Thus we may assume that f has at most one zero in $\mathbb{C}$. If f is a polynomial, this shows that $f(z) = A(z - B)^n$ for some $n \geqslant 0$. Assume then that f has an essential singularity at infinity. Let $a_n \to \infty$ be such that $f(a_n) \to 0$, and such that $a_n/|a_n| \to s$ for some $|s| = 1$. If $f(z_0) = 0$ for some z_0, then the point

$$b_n = z_0 + (a_n - z_0)/|a_n - z_0|$$

lies on the segment $[z_0, a_n]$ and so $|f(b_n)| \leqslant |f(a_n)|$. Since $b_n \to z_0 + s$, this forces f to have

function on a ring R of sets is of bounded variation. Is the assertion true if R is an algebra of sets?

Solved by A. Brunnschweiler (Switzerland), F. S. Cater, A. Janssen (West Germany), R. Bruce Mericle, Charles Riley, J. C. Smit (Netherlands) and the proposers. All solvers gave an example where R is the ring of subsets A of a set X such that either A or $X - A$ is finite, defining $\mu(A)$ to be the cardinal of A for A finite, and minus the cardinal of $X - A$ for $X - A$ finite.

A Condition on Entire Functions

6279 [1979, 793; 1981, 542]. *Proposed by Lee A. Rubel, University of Illinois, Urbana-Champaign.*

Let $f(z)$ be an entire function such that the maximum modulus over every closed line segment L is achieved at one of the endpoints a and b of L; that is,

$$\max\{|f(z)|: z \in L\} = \max\{|f(a)|, |f(b)|\}.$$

Prove that $f(z)$ has either the form $A(z - B)^n$ or the form $A \exp Bz$, where A and B are constants and n is a nonnegative integer.

The proposer pointed out that there is a gap in the published solution. Below are two correct solutions.

I. *Solution by P. R. Chernoff and J. Essick, University of California, Berkeley.* If f has more than one zero, then $|f|$ vanishes on the segment joining two of its zeros; so f is identically zero. Assume henceforth that f is not identically zero.

Case 1. f has one zero. Without loss of generality we may suppose $f(0) = 0$. If f is a polynomial, it is necessarily of the form cz^n. But if f is not a polynomial, then by the Casorati-Weierstrass theorem there is a sequence of points p_n with $|p_n| \to \infty$ and $|f(p_n)| = \varepsilon_n \to 0$. By passing to a subsequence, we may suppose that the unit complex numbers $\omega_n = p_n/|p_n|$ converge to a limit ω. Then $|f(z)| \leqslant \varepsilon_n$ on the line segment joining 0 to p_n, so that $|f(z)|$ must be identically 0 on the ray through 0 along direction ω. This is a contradiction.

Case 2. f has no zeros. Then $f = e^g$ for an entire function g. The real part $u(z)$ of $g(z)$, when restricted to any line L, can have no local maxima. Consequently, if the directional derivative $D_\omega u(a)$ along the direction ω at the point a should be positive, it follows that $u(a + t\omega)$ must be nondecreasing for $t \geqslant 0$, and thus $u(z) \geqslant u(a)$ for all z on the ray emanating from a along direction ω. Note that $D_\omega u(a) = \operatorname{Re}(\omega g'(a))$.

We claim that g' is constant. If not, we can obviously find three points a_1, a_2, a_3 such that the numbers $c_k = g'(a_k)$ are very close to ω_0^k, $k = 1, 2, 3$, where $\omega_0 = e^{2\pi i/3}$. Suppose now that ω is any complex number with $\operatorname{Re}(\omega c_k) > 0$. Then $D_\omega u(a_k) > 0$, so that $u(z) \geqslant u(a_k)$ for all z in the half plane

$$P_k = \{z : \operatorname{Re}((z - a_k)c_k)\} \geqslant 0.$$

Now, by the choice of c_1, c_2, c_3, the union $P_1 \cup P_2 \cup P_3$ of these half planes includes all of $\mathbb{C}$ except possibly a bounded triangular region, and therefore $u(z)$ is bounded below on the entire plane. This means that $g(z)$ must be constant, contradicting our assumption about g'. Thus g' is a constant A, $g(z) = Az + B$, and $f(z) = \exp(Az + B)$.

II. *Solution by Jan Maly and Zdenek Vlasek, Charles University, Prague.* If f vanishes at the endpoints of some segment $L = [a, b]$, then $f \equiv 0$ on L, and therefore $f \equiv 0$ on $\mathbb{C}$. Thus we may assume that f has at most one zero in $\mathbb{C}$. If f is a polynomial, this shows that $f(z) = A(z - B)^n$ for some $n \geqslant 0$. Assume then that f has an essential singularity at infinity. Let $a_n \to \infty$ be such that $f(a_n) \to 0$, and such that $a_n/|a_n| \to s$ for some $|s| = 1$. If $f(z_0) = 0$ for some z_0, then the point

$$b_n = z_0 + (a_n - z_0)/|a_n - z_0|$$

lies on the segment $[z_0, a_n]$ and so $|f(b_n)| \leqslant |f(a_n)|$. Since $b_n \to z_0 + s$, this forces f to have

another zero $z_0 + s$. We may therefore assume that $f(z) \neq 0$. For sufficiently large n the point

$$c_n = z + (a_n - z)/(n|a_n - z|)$$

lies on the segment $[z, a_n]$, which implies that $|f(c_n)| \leq |f(z)|$. Clearly $c_n \to z$. Now

$$R\left(s\frac{f'(z)}{f(z)}\right) = \mathrm{Re}\left(\frac{1}{f(z)}\lim_n \frac{c_n - z}{|c_n - z|} \cdot \lim_n \frac{f(c_n) - f(z)}{c_n - z}\right)$$

$$= \lim_n \left(\frac{1}{|c_n - z|}\mathrm{Re}\left(\frac{f(c_n)}{f(z)} - 1\right)\right) \leq 0$$

because $\mathrm{Re}(f(c_n)/f(z)) \leq |f(c_n)/f(z)| \leq 1$. Since z was arbitrary, this shows that the range of f'/f is contained in a half-plane, consequently $f'/f \equiv$ constant. Thus in this case there are constants A, B such that $f(z) = A\exp(Bz)$.

Values Assumed by a Modular Polynomial

6340 [1981, 294]. *Proposed by J. M. Patin, St. Herblain, France, and H. Stark, University of California, San Diego.*

Given a nonconstant polynomial $f(x)$ over $\mathbb{Z}$ and a positive $m \in \mathbb{Z}$, show that there exists $n \in \mathbb{Z}$ and a prime $p \equiv 1 \pmod{m}$ such that p divides $f(n)$. (For a special case of this problem see Elementary Problem E 2883, page 483.)

Solution by John Cremona, University of Michigan. Without loss of generality assume that f is irreducible. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be the roots of f, let ω be a primitive mth root of unity, and put $K = \mathbb{Q}(\omega, \alpha_1, \alpha_2, \ldots, \alpha_k)$. Since K is normal over $\mathbb{Q}$, it follows from the Čebotarev density theorem (see *Algebraic Number Theory*, editors Cassels and Fröhlich, p. 227) that the asymptotic proportion of rational primes p which split completely in K is $1/[K:\mathbb{Q}]$.

Let p be such a prime, which does not divide the discriminant of K. Then p must split completely in any subfield of K, and in particular p splits completely in $\mathbb{Q}(\alpha_1)$. By a fundamental theorem of Dirichlet, since p does not divide the discriminant of $\mathbb{Q}(\alpha_1)$ and p splits completely in $\mathbb{Q}(\alpha_1)$, the congruence $f(x) \equiv 0 \pmod{p}$ has k distinct solutions.

Let Φ_m denote the mth cyclotomic polynomial. Arguing similarly, with $\mathbb{Q}(\alpha_1)$ replaced by $\mathbb{Q}(\omega)$, we see that the congruence $\Phi_m(x) \equiv 0 \pmod{p}$ has exactly $\phi(m)$ roots; let r be one of them. Then the order of r in $(\mathbb{Z}/p\mathbb{Z})^\times$ is exactly m, and hence m divides $p - 1$.

Thus p has the desired properties.

Also solved by Ronald Evans, Harold Stark, and Ernst Kleinert (W. Germany).

Lattices of Periodic Functions

6358 [1981, 623]. *Proposed by Simon Fitzpatrick and Lee Rubel, University of Illinois at Urbana-Champaign.*

The periodic functions on the real line $\mathbb{R}$ (of possibly different periods) form a partially ordered set under pointwise $\leq$. Do they form a lattice—that is, is there a sup and an inf ($\vee$ and $\wedge$) of any two periodic functions, not necessarily the pointwise sup and inf? How about the functions that are periodic and bounded? Periodic and continuous?

Solution by Alberto Facchini, Università di Padova, Italy. The answer is no in general and for the periodic bounded functions, and it is yes for the functions that are periodic and continuous.

Denote the supremum (infimum) of two functions in the partially ordered set by $\vee$ ($\wedge$) and the pointwise supremum (infimum) by sup (inf). If S is a subset of $\mathbb{R}$, let $\chi(S)$ be the characteristic function of S.

another zero $z_0 + s$. We may therefore assume that $f(z) \neq 0$. For sufficiently large n the point

$$c_n = z + (a_n - z)/(n|a_n - z|)$$

lies on the segment $[z, a_n]$, which implies that $|f(c_n)| \leqslant |f(z)|$. Clearly $c_n \to z$. Now

$$R\left(s\frac{f'(z)}{f(z)}\right) = \mathrm{Re}\left(\frac{1}{f(z)}\lim_n \frac{c_n - z}{|c_n - z|} \cdot \lim_n \frac{f(c_n) - f(z)}{c_n - z}\right)$$

$$= \lim_n \left(\frac{1}{|c_n - z|}\mathrm{Re}\left(\frac{f(c_n)}{f(z)} - 1\right)\right) \leqslant 0$$

because $\mathrm{Re}(f(c_n)/f(z)) \leqslant |f(c_n)/f(z)| \leqslant 1$. Since z was arbitrary, this shows that the range of f'/f is contained in a half-plane, consequently $f'/f \equiv$ constant. Thus in this case there are constants A, B such that $f(z) = A\exp(Bz)$.

Values Assumed by a Modular Polynomial

6340 [1981, 294]. *Proposed by J. M. Patin, St. Herblain, France, and H. Stark, University of California, San Diego.*

Given a nonconstant polynomial $f(x)$ over $\mathbb{Z}$ and a positive $m \in \mathbb{Z}$, show that there exists $n \in \mathbb{Z}$ and a prime $p \equiv 1 \pmod{m}$ such that p divides $f(n)$. (For a special case of this problem see Elementary Problem E 2883, page 483.)

Solution by John Cremona, University of Michigan. Without loss of generality assume that f is irreducible. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be the roots of f, let ω be a primitive mth root of unity, and put $K = \mathbb{Q}(\omega, \alpha_1, \alpha_2, \ldots, \alpha_k)$. Since K is normal over $\mathbb{Q}$, it follows from the Čebotarev density theorem (see *Algebraic Number Theory*, editors Cassels and Fröhlich, p. 227) that the asymptotic proportion of rational primes p which split completely in K is $1/[K:\mathbb{Q}]$.

Let p be such a prime, which does not divide the discriminant of K. Then p must split completely in any subfield of K, and in particular p splits completely in $\mathbb{Q}(\alpha_1)$. By a fundamental theorem of Dirichlet, since p does not divide the discriminant of $\mathbb{Q}(\alpha_1)$ and p splits completely in $\mathbb{Q}(\alpha_1)$, the congruence $f(x) \equiv 0 \pmod{p}$ has k distinct solutions.

Let Φ_m denote the mth cyclotomic polynomial. Arguing similarly, with $\mathbb{Q}(\alpha_1)$ replaced by $\mathbb{Q}(\omega)$, we see that the congruence $\Phi_m(x) \equiv 0 \pmod{p}$ has exactly $\phi(m)$ roots; let r be one of them. Then the order of r in $(\mathbb{Z}/p\mathbb{Z})^\times$ is exactly m, and hence m divides $p - 1$.

Thus p has the desired properties.

Also solved by Ronald Evans, Harold Stark, and Ernst Kleinert (W. Germany).

Lattices of Periodic Functions

6358 [1981, 623]. *Proposed by Simon Fitzpatrick and Lee Rubel, University of Illinois at Urbana-Champaign.*

The periodic functions on the real line $\mathbb{R}$ (of possibly different periods) form a partially ordered set under pointwise $\leqslant$. Do they form a lattice—that is, is there a sup and an inf ($\vee$ and $\wedge$) of any two periodic functions, not necessarily the pointwise sup and inf? How about the functions that are periodic and bounded? Periodic and continuous?

Solution by Alberto Facchini, Università di Padova, Italy. The answer is no in general and for the periodic bounded functions, and it is yes for the functions that are periodic and continuous.

Denote the supremum (infimum) of two functions in the partially ordered set by $\vee$ ($\wedge$) and the pointwise supremum (infimum) by sup (inf). If S is a subset of $\mathbb{R}$, let $\chi(S)$ be the characteristic function of S.

another zero $z_0 + s$. We may therefore assume that $f(z) \neq 0$. For sufficiently large n the point

$$c_n = z + (a_n - z)/(n|a_n - z|)$$

lies on the segment $[z, a_n]$, which implies that $|f(c_n)| \leqslant |f(z)|$. Clearly $c_n \to z$. Now

$$R\left(s\frac{f'(z)}{f(z)}\right) = \mathrm{Re}\left(\frac{1}{f(z)}\lim_n\frac{c_n - z}{|c_n - z|}\cdot\lim_n\frac{f(c_n) - f(z)}{c_n - z}\right)$$
$$= \lim_n\left(\frac{1}{|c_n - z|}\mathrm{Re}\left(\frac{f(c_n)}{f(z)} - 1\right)\right) \leqslant 0$$

because $\mathrm{Re}(f(c_n)/f(z)) \leqslant |f(c_n)/f(z)| \leqslant 1$. Since z was arbitrary, this shows that the range of f'/f is contained in a half-plane, consequently $f'/f \equiv$ constant. Thus in this case there are constants A, B such that $f(z) = A\exp(Bz)$.

Values Assumed by a Modular Polynomial

6340 [1981, 294]. *Proposed by J. M. Patin, St. Herblain, France, and H. Stark, University of California, San Diego.*

Given a nonconstant polynomial $f(x)$ over $\mathbb{Z}$ and a positive $m \in \mathbb{Z}$, show that there exists $n \in \mathbb{Z}$ and a prime $p \equiv 1 \pmod m$ such that p divides $f(n)$. (For a special case of this problem see Elementary Problem E 2883, page 483.)

Solution by John Cremona, University of Michigan. Without loss of generality assume that f is irreducible. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be the roots of f, let ω be a primitive mth root of unity, and put $K = \mathbb{Q}(\omega, \alpha_1, \alpha_2, \ldots, \alpha_k)$. Since K is normal over $\mathbb{Q}$, it follows from the Čebotarev density theorem (see *Algebraic Number Theory*, editors Cassels and Fröhlich, p. 227) that the asymptotic proportion of rational primes p which split completely in K is $1/[K:\mathbb{Q}]$.

Let p be such a prime, which does not divide the discriminant of K. Then p must split completely in any subfield of K, and in particular p splits completely in $\mathbb{Q}(\alpha_1)$. By a fundamental theorem of Dirichlet, since p does not divide the discriminant of $\mathbb{Q}(\alpha_1)$ and p splits completely in $\mathbb{Q}(\alpha_1)$, the congruence $f(x) \equiv 0 \pmod p$ has k distinct solutions.

Let Φ_m denote the mth cyclotomic polynomial. Arguing similarly, with $\mathbb{Q}(\alpha_1)$ replaced by $\mathbb{Q}(\omega)$, we see that the congruence $\Phi_m(x) \equiv 0 \pmod p$ has exactly $\phi(m)$ roots; let r be one of them. Then the order of r in $(\mathbb{Z}/p\mathbb{Z})^\times$ is exactly m, and hence m divides $p - 1$.

Thus p has the desired properties.

Also solved by Ronald Evans, Harold Stark, and Ernst Kleinert (W. Germany).

Lattices of Periodic Functions

6358 [1981, 623]. *Proposed by Simon Fitzpatrick and Lee Rubel, University of Illinois at Urbana-Champaign.*

The periodic functions on the real line $\mathbb{R}$ (of possibly different periods) form a partially ordered set under pointwise $\leqslant$. Do they form a lattice—that is, is there a sup and an inf ($\vee$ and $\wedge$) of any two periodic functions, not necessarily the pointwise sup and inf? How about the functions that are periodic and bounded? Periodic and continuous?

Solution by Alberto Facchini, Università di Padova, Italy. The answer is no in general and for the periodic bounded functions, and it is yes for the functions that are periodic and continuous.

Denote the supremum (infimum) of two functions in the partially ordered set by $\vee$ ($\wedge$) and the pointwise supremum (infimum) by sup (inf). If S is a subset of $\mathbb{R}$, let $\chi(S)$ be the characteristic function of S.

To show that the answer is no in general and for the functions that are periodic and bounded, we consider the functions

$$f_1 = \chi(\mathbb{Z}), \quad f_2 = \chi(\mathbb{Z} \cdot \sqrt{2}), \quad f_3 = \chi(\mathbb{Z} + \mathbb{Z} \cdot \sqrt{2}),$$
$$f_4 = \chi((\mathbb{Z} + \mathbb{Z} \cdot \sqrt{3}) \cup (\mathbb{Z} \cdot \sqrt{2} + \mathbb{Z} \cdot \sqrt{3})), \quad f_5 = \chi(\mathbb{Z} \cup \mathbb{Z} \cdot \sqrt{2}).$$

Then f_1, f_2, f_3, f_4 are periodic with periods $1, \sqrt{2}, 1, \sqrt{3}$, respectively, and f_5 is not periodic. If $f_1 \vee f_2$ exists, then

$$f_5 = \sup(f_1, f_2) \leqslant f_1 \vee f_2 \leqslant \inf(f_3, f_4) = f_5,$$

i.e., $f_1 \vee f_2 = f_5$ is not periodic, and this is a contradiction.

To show that the functions that are periodic and continuous do form a lattice, it is obviously enough to prove that any two such functions have a $\vee$. Let f_1, f_2 be continuous periodic functions with A and B as periods respectively. If A/B is rational, $A/B = m/n$ say, then $f_1 \vee f_2 = \sup(f_1, f_2)$ has $nA = mB$ as a period. Suppose A/B is not rational. Let $M_i = \max f_i$ for $i = 1, 2$ (M_i exists because f_i is periodic and continuous). We may assume $M_1 \leqslant M_2$. Let $g = \sup(M_1, f_2)$. We show that g (continuous and periodic of period B) is $f_1 \vee f_2$. Let h be any periodic continuous function of period C, such that $h \geqslant f_1, f_2$. Then either A/C or B/C is not rational. If A/C is not rational and $f_1(x_0) = M_1$, then, for all integers n, m, we have

$$h(x_0 + nA + mC) = h(x_0 + nA) \geqslant f_1(x_0 + nA) = f_1(x_0) = M_1.$$

The set $\{x_0 + nA + mC | n, m \in \mathbb{Z}\}$ is dense in $\mathbb{R}$ and h is continuous; therefore $h \geqslant M_1$ and $h \geqslant g$. Similarly if B/C is not rational, then $h \geqslant M_2$ and therefore $h \geqslant \sup(M_1, f_2) = g$. This proves that $g = f_1 \vee f_2$.

Also solved by F. S. Cater, James Caveny, G. A. Edgar, O. P. Lossers (The Netherlands), Mark D. Meyerson, and the proposers. Partially solved by Bruce Richter and Herb Shank (Canada).

The Distribution Function of $\phi(n)/n$

6363 [1981, 710–711]. *Proposed by H. G. Diamond, University of Illinois and P. Erdős, Hungarian Academy of Sciences.*

Let ϕ denote Euler's function and set

$$D(\alpha) = \lim_{x \to \infty} \frac{1}{x} \#\{n \leqslant x : \phi(n)/n \leqslant \alpha\},$$

the distribution function of $\phi(n)/n$. Prove that

$$D\left(\frac{1}{2}\right) - D\left(\frac{1}{4}\right) + D\left(\frac{1}{8}\right) - D\left(\frac{1}{16}\right) + \cdots = \frac{1}{2}.$$

Solution by O. P. Lossers, Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, the Netherlands. Define, for $l \geqslant 0$,

$$D_l(\alpha, x) = \frac{1}{x} \#\{n \leqslant x : n = 2^l m, \quad m \text{ odd and } \phi(n)/n \leqslant \alpha\},$$
$$D_l(\alpha) = \lim_{x \to \infty} D_\alpha(\alpha, x).$$

The existence of $D_l(\alpha)$ can be established in much the same way as the existence of $D(\alpha)$ was established by I. J. Schoenberg, Über die Asymptotische Verleitung Reeler Zahlen mod 1, Math. Z., 28 (1928) 171–199. Evidently

(1) $$\frac{1}{x} \#\{n \leqslant x : \phi(n)/n \leqslant \alpha\} = \sum_{l=0}^{\infty} D_l(\alpha, x).$$

To show that the answer is no in general and for the functions that are periodic and bounded, we consider the functions

$$f_1 = \chi(\mathbb{Z}), \quad f_2 = \chi(\mathbb{Z}\cdot\sqrt{2}), \quad f_3 = \chi(\mathbb{Z} + \mathbb{Z}\cdot\sqrt{2}),$$
$$f_4 = \chi((\mathbb{Z} + \mathbb{Z}\cdot\sqrt{3}) \cup (\mathbb{Z}\cdot\sqrt{2} + \mathbb{Z}\cdot\sqrt{3})), \quad f_5 = \chi(\mathbb{Z} \cup \mathbb{Z}\cdot\sqrt{2}).$$

Then f_1, f_2, f_3, f_4 are periodic with periods $1, \sqrt{2}, 1, \sqrt{3}$, respectively, and f_5 is not periodic. If $f_1 \vee f_2$ exists, then

$$f_5 = \sup(f_1, f_2) \leqslant f_1 \vee f_2 \leqslant \inf(f_3, f_4) = f_5,$$

i.e., $f_1 \vee f_2 = f_5$ is not periodic, and this is a contradiction.

To show that the functions that are periodic and continuous do form a lattice, it is obviously enough to prove that any two such functions have a $\vee$. Let f_1, f_2 be continuous periodic functions with A and B as periods respectively. If A/B is rational, $A/B = m/n$ say, then $f_1 \vee f_2 = \sup(f_1, f_2)$ has $nA = mB$ as a period. Suppose A/B is not rational. Let $M_i = \max f_i$ for $i = 1, 2$ (M_i exists because f_i is periodic and continuous). We may assume $M_1 \leqslant M_2$. Let $g = \sup(M_1, f_2)$. We show that g (continuous and periodic of period B) is $f_1 \vee f_2$. Let h be any periodic continuous function of period C, such that $h \geqslant f_1, f_2$. Then either A/C or B/C is not rational. If A/C is not rational and $f_1(x_0) = M_1$, then, for all integers n, m, we have

$$h(x_0 + nA + mC) = h(x_0 + nA) \geqslant f_1(x_0 + nA) = f_1(x_0) = M_1.$$

The set $\{x_0 + nA + mC \mid n, m \in \mathbb{Z}\}$ is dense in $\mathbb{R}$ and h is continuous; therefore $h \geqslant M_1$ and $h \geqslant g$. Similarly if B/C is not rational, then $h \geqslant M_2$ and therefore $h \geqslant \sup(M_1, f_2) = g$. This proves that $g = f_1 \vee f_2$.

Also solved by F. S. Cater, James Caveny, G. A. Edgar, O. P. Lossers (The Netherlands), Mark D. Meyerson, and the proposers. Partially solved by Bruce Richter and Herb Shank (Canada).

The Distribution Function of $\phi(n)/n$

6363 [1981, 710–711]. *Proposed by H. G. Diamond, University of Illinois and P. Erdős, Hungarian Academy of Sciences.*

Let ϕ denote Euler's function and set

$$D(\alpha) = \lim_{x\to\infty} \frac{1}{x}\#\{n \leqslant x : \phi(n)/n \leqslant \alpha\},$$

the distribution function of $\phi(n)/n$. Prove that

$$D\left(\frac{1}{2}\right) - D\left(\frac{1}{4}\right) + D\left(\frac{1}{8}\right) - D\left(\frac{1}{16}\right) + \cdots = \frac{1}{2}.$$

Solution by O. P. Lossers, Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, the Netherlands. Define, for $l \geqslant 0$,

$$D_l(\alpha, x) = \frac{1}{x}\#\{n \leqslant x : n = 2^l m, \quad m \text{ odd and } \phi(n)/n \leqslant \alpha\},$$
$$D_l(\alpha) = \lim_{x\to\infty} D_\alpha(\alpha, x).$$

The existence of $D_l(\alpha)$ can be established in much the same way as the existence of $D(\alpha)$ was established by I. J. Schoenberg, Über die Asymptotische Verleitung Reeler Zahlen mod 1, Math. Z., 28 (1928) 171–199. Evidently

$$(1) \qquad \frac{1}{x}\#\{n \leqslant x : \phi(n)/n \leqslant \alpha\} = \sum_{l=0}^{\infty} D_l(\alpha, x).$$

For $n = 2^l m$, m odd, $l \geqslant 1$, we have $\phi(n)/n = \phi(m)/(2m)$, and hence

$$D_l(\alpha, x) = 2^{-l} D_0(2\alpha, x) \quad \text{for } l \geqslant 1. \tag{2}$$

Since $D_0(2\alpha, x) \leqslant 1$, it follows from (1) and (2), by the Weierstrass M-test, that

$$D(\alpha) = D_0(\alpha) + \sum_{l=1}^{\infty} 2^{-l} D_0(2\alpha) = D_0(\alpha) + D_0(2\alpha).$$

Consequently,

$$\sum_{k=1}^{n} (-1)^{k-1} D(2^{-k}\alpha) = D_0(\alpha) + (-1)^{n-1} D_0(2^{-n}\alpha) \to D_0(\alpha)$$

as $n \to \infty$, the distribution function D_0 being upper semicontinuous at 0. Hence

$$D\left(\frac{\alpha}{2}\right) - D\left(\frac{\alpha}{4}\right) + D\left(\frac{\alpha}{8}\right) - D\left(\frac{\alpha}{16}\right) + \cdots = D_0(\alpha)$$

and this equals 1 when $\alpha \geqslant 1$.

Also solved by Robert Breusch, William A. Newcomb, R. K. Odoni (U.K.), Lajos Takács, Charles R. Wall, and the proposers.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

A History of the Calculus of Variations from the 17th through the 19th Century. By Herman H. Goldstine. Studies in the History of Mathematics and Physical Sciences. Edited by G. J. Toomer, vol. 5. Springer-Verlag, New York/Heidelberg/Berlin, 1980. pp. xviii + 410, ISBN 0-387-90521-9.

HELENA M. PYCIOR
Department of History, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

There are two kinds of history of mathematics. One, written primarily by mathematicians, examines the past from a presentist perspective. It seeks the roots of modern mathematics in the answers to such questions as: When was this result first proved, and by whom? Was the original proof correct? This kind of history is usually "technical...[and] written on a high mathematical level" [**2**, p. 439]. The other, written primarily by historians, evidences more concern for the past per se. This second kind of history tends to be less technical, more expository, and broader in perspective. It sees historical periods as rich fabrics, and particular mathematical developments as threads in those fabrics. Thus, in this kind of history, study of a particular development becomes, to a reasonable extent, study of the period in which it occurred. As far as they relate to the given topic, the period's general mathematics, and possibly even science, philosophy, religion, politics, and education are considered.

The book under review is clearly a mathematician's history, evidencing the strengths and some of the weaknesses of that historical genre. It details effectively evolution of the calculus of variations from its roots in Fermat's papers of 1662 on the passage of light through optical media to Hilbert's discovery of the invariant integral and existence theory. At the same time, it highlights the importance of problems (as stimuli), of generalization (as method), and of physical applicability (as incentive) to the development of the calculus of variations. These three themes appear early

For $n = 2^l m$, m odd, $l \geqslant 1$, we have $\phi(n)/n = \phi(m)/(2m)$, and hence

$$D_l(\alpha, x) = 2^{-l} D_0(2\alpha, x) \quad \text{for } l \geqslant 1. \tag{2}$$

Since $D_0(2\alpha, x) \leqslant 1$, it follows from (1) and (2), by the Weierstrass M-test, that

$$D(\alpha) = D_0(\alpha) + \sum_{l=1}^{\infty} 2^{-l} D_0(2\alpha) = D_0(\alpha) + D_0(2\alpha).$$

Consequently,

$$\sum_{k=1}^{n} (-1)^{k-1} D(2^{-k}\alpha) = D_0(\alpha) + (-1)^{n-1} D_0(2^{-n}\alpha) \to D_0(\alpha)$$

as $n \to \infty$, the distribution function D_0 being upper semicontinuous at 0. Hence

$$D\left(\frac{\alpha}{2}\right) - D\left(\frac{\alpha}{4}\right) + D\left(\frac{\alpha}{8}\right) - D\left(\frac{\alpha}{16}\right) + \cdots = D_0(\alpha)$$

and this equals 1 when $\alpha \geqslant 1$.

Also solved by Robert Breusch, William A. Newcomb, R. K. Odoni (U.K.), Lajos Takács, Charles R. Wall, and the proposers.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

A History of the Calculus of Variations from the 17th through the 19th Century. By Herman H. Goldstine. Studies in the History of Mathematics and Physical Sciences. Edited by G. J. Toomer, vol. 5. Springer-Verlag, New York/Heidelberg/Berlin, 1980. pp. xviii + 410, ISBN 0-387-90521-9.

HELENA M. PYCIOR
Department of History, University of Wisconsin-Milwaukee, Milwaukee, WI 53201

There are two kinds of history of mathematics. One, written primarily by mathematicians, examines the past from a presentist perspective. It seeks the roots of modern mathematics in the answers to such questions as: When was this result first proved, and by whom? Was the original proof correct? This kind of history is usually "technical...[and] written on a high mathematical level" [**2**, p. 439]. The other, written primarily by historians, evidences more concern for the past per se. This second kind of history tends to be less technical, more expository, and broader in perspective. It sees historical periods as rich fabrics, and particular mathematical developments as threads in those fabrics. Thus, in this kind of history, study of a particular development becomes, to a reasonable extent, study of the period in which it occurred. As far as they relate to the given topic, the period's general mathematics, and possibly even science, philosophy, religion, politics, and education are considered.

The book under review is clearly a mathematician's history, evidencing the strengths and some of the weaknesses of that historical genre. It details effectively evolution of the calculus of variations from its roots in Fermat's papers of 1662 on the passage of light through optical media to Hilbert's discovery of the invariant integral and existence theory. At the same time, it highlights the importance of problems (as stimuli), of generalization (as method), and of physical applicability (as incentive) to the development of the calculus of variations. These three themes appear early

(but implicitly) in the book. The first chapter, for example, deals with Newton's problem of motion in a resisting medium, and the brachystochrone problem. Newton's problem is to find the solid of revolution which moves through a resisting medium in the direction of its axis with the least resistance. Goldstine describes the problem as "the first genuine problem of the calculus of variations" (p. 7), partially because it was the first in the field to be correctly solved, and because techniques used in its solution proved useful in the later development of the field. Newton's techniques were used, in particular, in some solutions of the brachystochrone problem. Solved early by Newton, Leibniz, and the Bernoulli brothers, this problem is: given two points in a vertical plane, find the curve joining those points along which a frictionless particle will fall in the shortest time. Goldstine's emphasis is on James Bernoulli's solution which, although less elegant than his brother's, was more general and therefore applicable to other problems. These additional problems, together with Newton's and the brachystochrone, constituted a new field of mathematics.

Chapter 2 covers the work of Euler, who was probably led into the field by the Bernoullis and by 1744 gave it a general framework. Euler also enunciated the principle of least action, when he wrote: "all the effects of Nature follow a certain law of maxima or minima" (p. 106). Discussed in some detail by Goldstine, this principle, of course, became the field's physical raison d'être. Euler's general approach to the calculus of maxima and minima was dependent on geometry through 1755. In that year, Lagrange (then only nineteen) proposed the idea of a variation which enabled mathematicians to bypass the geometrical aspects of Euler's theory and to develop a truly analytic theory in its place. Euler seized Lagrange's idea and renamed the field of their joint endeavors the calculus of variations. Later chapters, covering Jacobi, Weierstrass, Clebsch, Mayer, Hilbert, and Kneser, continue the story of the emergence of the classical theory of calculus of variations (which, Goldstine maintains, was essentially completed by Hilbert's work) and further underscore the vital roles of problems, generalization, and physical applicability in that theory.

Goldstine's deep understanding of the calculus of variations inspires his history and assures its mathematical accuracy. Thus, as a mathematician, he may be trusted "to select [judiciously] those papers and authors whose works have played key roles in the classical calculus of variations as we understand the subject today" (p. vii). Furthermore, when Goldstine explains the content of these select works, his mathematical ability enables him to translate readily from old concepts and notation to that of modern mathematics. In his discussion of Newton's problem of motion in a resisting medium, for example, he gives Newton's solution in more modern notation first, and only later in the original. This technique clearly facilitates the reader's transition from the familiar to the archaic.

Goldstine's experience as a mathematician positively influences not only his selection and explanation of historical texts, but the questions he puts to the texts as well. Thus the reader feels the touch of the mathematician as Goldstine deals with the concepts of the calculus of variations on two levels—as "formalistic constructs" and as ideas with some physical or geometrical backing. This bilevel analysis marks, for example, the discussion of the relative contributions of Euler and Lagrange to the idea of a variation. Relying apparently on little firm evidence but rather on mathematical instinct, Goldstine concludes that Lagrange probably viewed his variation operator "as a purely formalistic construct" (p. 112), and that Euler was the first to see it as a means of effecting a comparison of curves.

Yet, in spite of Goldstine's many strengths as a mathematician-historian, parts of his book contain historical shortcomings. Sometimes Goldstine is guilty of extreme presentism; other times, of hero worship and of tunnel vision. Goldstine's presentist perspective leads him to bypass what Kenneth O. May, the late dean of North American historians of mathematics, once described as "understanding past work as it *was*, rather than as we would *now* do it" [**3**, p. 451]. In the section on Leibniz's solution of the brachystochrone problem, for example, Goldstine dismisses part of Leibniz's proof as "unnecessary" and goes directly to what he considers its "essence." He gives no indication of the nature of the superfluous material, or of the possible reasons behind its inclusion

in the original proof. Goldstine's treatment of Clebsch's work on the problem of Lagrange is even more objectionable: here he does not merely delete original material, but actually replaces it with his own. Complementing an original proof with a modern formulation (as in Goldstine's treatment of Newton) is laudable; omitting part of the original (as with Leibniz) is tolerable. But substituting for the original is inexcusable. In his handling of Clebsch's work, then, Goldstine crosses the line from history to mathematics. Abandoning the historian's task of recreating the past, Goldstine redoes it.

At least once in the book, Goldstine also crosses the line between appropriate respect for past mathematical genius and hero worship. "The historian," May once wrote, "finds hilarious such naive historical mistakes as assuming that... a brilliant mathematician of past centuries must have understood a concept or had a proof because these would be evident to lesser lights today" [**3**, p. 453]. This mistake is precisely the kind of hero worship to which Goldstine falls victim. He laments Newton's failure to include a key observation in his lean solution of the brachystochrone problem. Then he says that Newton knew what he failed to record. "It is reasonable," Goldstine argues, "that he must have seen [this] at once... That is, he certainly must have been aware of the fact... ." (p. 35). The historical fact is that Newton did not record the observation, and really that is all that Goldstine can or should state at this point.

More substantively, in its quest for mathematical detail, Goldstine's work sacrifices breadth of coverage. True to the genre of the mathematician's history of mathematics, Goldstine very narrowly delimits his topic. Replete with equations and sometimes multiple proofs of the same theorem, his work evidences little interest in possible external influences on the development of the calculus of variations, or even a wide mathematical perspective on its subject. Yet, these more general considerations, as the book itself suggests, are not artificial, but arise naturally in discussion of the history of the calculus of variations. In his section on priority for the principle of least action, Goldstine notes that Carathéodory ascribed a philosophical origin to Euler's statement of the principle. Evidencing historical tunnel vision, Goldstine pursues this issue no further, leaving the reader with but a footnote to Carathéodory's work. If a broader extrinsic perspective seems advisable, a perspective which takes into account the general development of the differential and integral calculus seems mandatory. Yet Goldstine actually says very little about the calculus. In fact, he misses a few golden opportunities to connect its development with that of the calculus of variations. Thus, for example, he fails to study Lagrange's work in the context of the eighteenth-century movement in the calculus away from geometrical and towards analytical foundations.

There is, finally, one general criticism of the book. Its style is not representative of the finest mathematical (or historical) exposition. It contains no adequate statement of purpose or audience, makes but few attempts to highlight major ideas through clear prose statements, and evidences throughout an extremely tight, technical style. The history of mathematics is not mathematics, and seemingly should be directed to audiences beyond mathematical specialists—to general historians of science and mathematics majors, in particular. Admittedly, the nature of the calculus of variations probably precludes reduction of its history to prose statements and easily understandable equations. But the expository shortcomings of Goldstine's book cannot all be attributed to the mathematical complexity of its subject. The first volume of the Carus Mathematical Monographs series was *Calculus of Variations* [**1**], Gilbert Ames Bliss's eminently readable account of the field.

When one puts aside speculations concerning how much Goldstine's book could have been improved had he avoided some common historical pitfalls and adopted a less technical writing style, the conclusion is that the work is a solid contribution written in the tradition of the mathematician's history of mathematics. It will probably soon emerge as a standard source on the history of the calculus of variations, and be extended by representatives of both genres of the history of mathematics. But, clearly, responsibility for further development of the calculus of variations now rests with practitioners of the second kind of history of mathematics. Historians

should begin with Goldstine's work and add to it historical texture—biographical detail and exploration of the past interrelationships among the calculus of variations, general mathematics, science, and appropriate cultural factors such as philosophy.

References

1. Gilbert Ames Bliss, Calculus of Variations, Carus Mathematical Monograph No. 1, Mathematical Association of America, Washington, D.C., 1925.

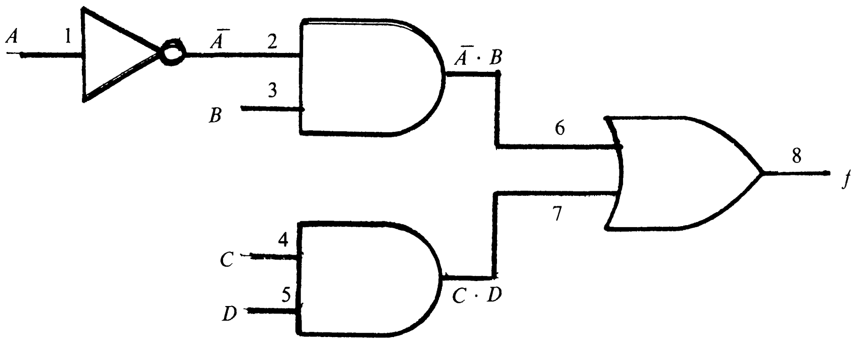
2. Judith V. Grabiner, The mathematician, the historian, and the history of mathematics, Historia Math., 2 (1975) 439–447.

3. Kenneth O. May, What is good history and who should do it?, Historia Math., 2 (1975) 449–455.

Computer Logic, Testing and Verification. By Paul Roth. Pitman Publishing Limited, London, 1980. xx + 176 pp.

J. R. ARMSTRONG
Electrical Engineering Department, Virginia Polytechnic Institute, Blacksburg, VA 24061

1. Introduction. Like all man-made devices, any computer that is built will eventually fail. It is important that these failures be detected so that the computer's results will not be accepted as being computationally correct while it is in the failed state. The problem of fault detection in computer circuits is thus an important research area for computer scientists and engineers [**1**], [**2**].



A Combinational Logic Circuit.

FIG. 1

Computer circuitry is of two basic types: combinational logic and sequential logic. Combinational logic is made up of gates which in their most primitive form implement the logical AND, OR, and INVERT functions. A combinational logic function is thus composed from these basic primitives; for example, the logic circuit shown in Fig. 1 has the output function:

$$f = (\bar{A} \cdot B) + (C \cdot D).$$

Combinational logic circuits have no memory; i.e., the value of the output is solely dependent upon the value of the inputs, and thus will change only when the inputs change and the circuit function dictates an output change. In real logic circuits, of course, there is a signal propagation delay between input and output changes.

Sequential logic has memory. The value of the output at a future time is dependent upon the present state of the device as well as the present value of its inputs. Fig. 2 shows the state table for

should begin with Goldstine's work and add to it historical texture—biographical detail and exploration of the past interrelationships among the calculus of variations, general mathematics, science, and appropriate cultural factors such as philosophy.

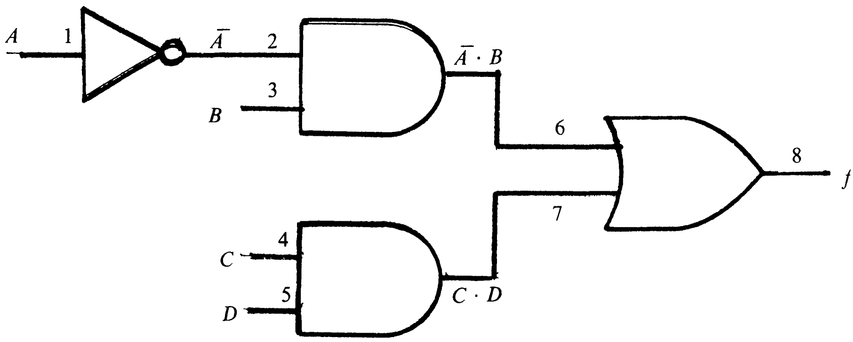
References

1. Gilbert Ames Bliss, Calculus of Variations, Carus Mathematical Monograph No. 1, Mathematical Association of America, Washington, D.C., 1925.
2. Judith V. Grabiner, The mathematician, the historian, and the history of mathematics, Historia Math., 2 (1975) 439–447.
3. Kenneth O. May, What is good history and who should do it?, Historia Math., 2 (1975) 449–455.

Computer Logic, Testing and Verification. By Paul Roth. Pitman Publishing Limited, London, 1980. xx + 176 pp.

J. R. ARMSTRONG
Electrical Engineering Department, Virginia Polytechnic Institute, Blacksburg, VA 24061

1. Introduction. Like all man-made devices, any computer that is built will eventually fail. It is important that these failures be detected so that the computer's results will not be accepted as being computationally correct while it is in the failed state. The problem of fault detection in computer circuits is thus an important research area for computer scientists and engineers [1], [2].



A Combinational Logic Circuit.

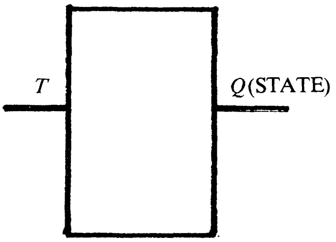
FIG. 1

Computer circuitry is of two basic types: combinational logic and sequential logic. Combinational logic is made up of gates which in their most primitive form implement the logical AND, OR, and INVERT functions. A combinational logic function is thus composed from these basic primitives; for example, the logic circuit shown in Fig. 1 has the output function:

$$f = (\bar{A} \cdot B) + (C \cdot D).$$

Combinational logic circuits have no memory; i.e., the value of the output is solely dependent upon the value of the inputs, and thus will change only when the inputs change and the circuit function dictates an output change. In real logic circuits, of course, there is a signal propagation delay between input and output changes.

Sequential logic has memory. The value of the output at a future time is dependent upon the present state of the device as well as the present value of its inputs. Fig. 2 shows the state table for



T	Old State	New State
0	0	0
0	1	1
1	0	1
1	1	0

Toggle (T) Flipflop.

FIG. 2

a "toggle (T) flipflop." Note that the device retains its old state when $T = 0$, but inverts its old state when $T = 1$. In sequential circuits, the state to state transition times are controlled either by internal circuit signal delays or by the period of a clock which may be applied to the element.

In real computer circuits combinational and sequential logic are frequently mixed within a given logic card, but the approach to testing the two types of logic is different. Let us first consider tests for combinational logic.

In developing these tests, one usually assumes that only one fault is present. Also, the fault model that is frequently employed is the "stuck-at" model, i.e., the effect of the fault is to stick a signal line at either 1 or at 0. Thus only permanent faults are considered with this model. Note in Fig. 1 that all of the signal lines have been numbered. Assume that the fault, line 1 stuck at a 0, is present. To develop a test for this fault, we determine an input vector (A, B, C, D) which will give one output when the fault is present and the logical complement of this output when it is not. There are three basic approaches to this test development: intuitive (which we will not consider here), algebraic, and algorithmic.

First let us consider an example of an algebraic technique. The first step is to derive the output function for the circuit in the presence of the fault. For the circuit in Fig. 1 this function is:

$$f' = B + C \cdot D.$$

We next employ the EXCLUSIVE OR operation ($\oplus$) which has the following definition:

$$x \oplus y = x \cdot \bar{y} + \bar{x} \cdot y.$$

The set of tests (T) for the fault in question is then given by the expression:

$$T = f \oplus f'; = (\bar{A} \cdot B + C \cdot D) \cdot \overline{(B + C \cdot D)} + \overline{(\bar{A} \cdot B + C \cdot D)} \cdot (B + C \cdot D),$$

which can be simplified using postulates of Boolean Algebra to the form:

$$T = AB(\bar{C} + \bar{D}).$$

T gives us the complete set of tests for fault, in this case: $A = 1, B = 1, C = 0, D = X$ (don't care) and $A = 1$, $B = 1$, $C = X$, $D = 0$. Algebraic methods of this type will always work, in theory, provided that the test exists. However, for circuits of even reasonable size the initial expressions for T become very complex and the simplification of these expressions using the postulates of Boolean algebra is a difficult process not amenable to automation.

The most important algorithmic approach to logic testing is the "D algorithm." It is based on a calculus of five values: $0, 1, X, D$ and $\bar{D}$. 0 and 1 have their normal meaning, X denotes a logic

signal value which is unspecified, D a logic signal value which has the value 0 when the fault is present and 1 otherwise, and $\bar{D}$ a logic signal value that has the value 1 when the fault is present and 0 otherwise. Using these five values, the state of an n wire logic circuit is then represented as an n dimensional vector (cube). Fig. 3 shows two such cubes for the circuit in Fig. 1.

LINES:	1	2	3	4	5	6	7	8
CUBES:	D	$\bar{D}$	1	0	X	$\bar{D}$	0	$\bar{D}$
	D	$\bar{D}$	1	X	0	$\bar{D}$	0	$\bar{D}$

Test Cubes for the Circuit in Fig. 1.

FIG. 3

The two cubes shown in Fig. 3 are in fact "test cubes" for the fault, line 1 stuck at a 0. Note that the output wire has the value $\bar{D}$ and thus will have the value 1 if this fault is present, the value 0 otherwise. The values on lines 1, 2 and 6 also switch appropriately depending on the presence of this fault. The two test cubes are in fact the end product of the D algorithm which assigns the line at the site of the fault a value D for a stuck at 0 and a value $\bar{D}$ for a stuck at 1. It then determines the conditions necessary to propagate the D or $\bar{D}$ to the circuit output. These conditions are then "justified" in terms of input signal values, thus developing a test for the fault. In order to propagate the D or $\bar{D}$ to the output, it is necessary to redefine the AND, OR and INVERT to allow for D and $\bar{D}$ inputs (see FIG. 4 below).

+	0 1 D $\bar{D}$		0 1 D $\bar{D}$		Inv
0	0 1 D $\bar{D}$	0	0 0 0 0	0	1
1	1 1 1 1	1	0 1 D $\bar{D}$	1	0
D	D 1 D 1	D	0 D D 0	D	$\bar{D}$
$\bar{D}$	$\bar{D}$ 1 1 $\bar{D}$	$\bar{D}$	0 $\bar{D}$ 0 $\bar{D}$	$\bar{D}$	D

Primitives for Propagation of D and $\bar{D}$.

FIG. 4

Thus for our example, in order to drive the D on line 1 to the output, the algorithm first determines that line 3 has to be a logic 1 and that line 7 has to be a logic 0. The 0 on line 7 is then justified by setting either line 4 or line 5 to 0, thus generating the two test cubes for the fault, line 1 stuck at a 0.

Based on our interpretation of the values D and $\bar{D}$, note that the above two tests also detect the faults lines 2, 6 and 8 stuck at a 1. Although not evident from Fig. 3, further reflection will show that the two tests also detect the fault line 7, stuck at a 1. A typical approach then is to: (1) assume the existence of a stuck-at-fault, (2) use the D algorithm to determine test(s) for that fault and as a by-product other faults along the drive path, (3) use "further reflection," as implemented in another algorithm, to determine other faults detected by the tests determined in (2). This approach has been found to be very effective and is widely used in the computer industry today.

When a logic circuit contains sequential logic, i.e., has memory, the testing problem is greatly complicated by the fact that the states of these devices are unknown in the presence of faults. Because of this, from a theoretical point of view, the testing problem has never been satisfactorily solved for sequential circuits. An approach that has worked successfully in practice is to provide special test inputs and outputs which can be used to preset and observe the states of memory elements. Once this capability is in place, the rest of the circuit, being combinational logic, can be tested using the testing techniques described above. For example, suppose that the circuit in Fig. 1 receives its inputs (A, B, C, D), not from system input lines, but from four memory devices (flipflops). Then if one has the ability to provide inputs to these flipflops (Scan In) and observe that the flipflops have been set correctly (Scan Out), the rest of the circuit is easily tested.

The word "authority" is much misused, but if any one deserves this characterization in the field of logic testing, it is Paul Roth, author of Computer Logic, Testing and Verification. His experience in the field of logic design and testing dates back to his early work with the Von Neumann team at the Institute for Advanced Study, which built the first stored program Computer in the United States. However, his most significant accomplishment was the development of cubical calculus and its most important application in the field of logic testing: the D algorithm. In addition he has been instrumental in the development of effective means of "further reflection" for determining all faults covered by a test input. This work at IBM has been a major factor in the development of a test capability that has developed tests for systems as large as 750,000 gates!

The cubical calculus that Roth developed has other important applications. It provides a basic method of circuit description which can be used for logic minimization and verification. Logic minimization involves finding a minimal set of cubes which cover the function. Fig. 5 below gives a minimal "cover" for the function implemented by the circuit in Fig. 1.

Line No.	1	2	3	4	5	6	7	8
	0	1	1	X	X	1	X	1
	X	X	X	1	1	X	1	1

Minimal Cover for the Circuit in Fig. 1.

FIG. 5

If one has the cubical description for two circuits, these descriptions can be compared to determine if the two circuits perform the same function. Thus it can also be used for design verification, both in initial phases of the design process, and also during latter phases when engineering changes are made to the logic. In summary, cubical calculus provides a basis for the algorithmic solution of three of the most important problems in the field of design automation.

One of the ironies of this LSI era of logic design is that, even though the capability of being able to put tens of thousands of gates on a single chip has allowed us to design very reliable systems, the problem of testing these devices has become progressively more difficult as the chip size increases. This is because the observability of internal logic signals decreases with chip size. Also, in order to place more gates on a single chip, it is necessary to shrink the physical size of the individual gates and also the spacing between interconnect lines. Thus if one has a physical defect in a chip, it is increasingly likely that it will cause a number of gates to malfunction. This brings into question the single fault assumption mentioned above and also the stuck-at-fault model which is the basis for most approaches to logic testing. It is increasingly important to develop test techniques which treat an LSI device as a single functional entity instead of a collection of thousands of gates. To be sure, various approaches to functional level modeling are beginning to appear [3] but corresponding activity in test algorithm design using higher level models has lagged. Some of this is due to the complexity of the problem, but it also may be that the computer industry has gotten too comfortable with the gate level model. The challenge then is to develop high level modeling techniques which will not only reproduce the behavior of good LSI devices but faulty ones as well. A particularly interesting problem is that of determining how physical defects in integrated circuits map into effects at the functional model level. Solutions to problems of this type are necessary if advances in logic testing techniques are to keep pace with the technological explosion that is VLSI.

References

1. A. D. Friedman and F. P. Menon, Fault Detection in Digital Circuits, Prentice-Hall, Englewood Cliffs, NJ, 1971.

2. M. A. Breuer and A. D. Friedman, Diagnosis and Design of Digital Systems, Computer Science Press, Woodland Hills, CA, 1976.

3. J. R. Armstrong and D. E. Devlin, GSP, A Logic Simulator for LSI, Proceedings of the 18th Design Automation Conference, Nashville, TN (Sponsored by the ACM and IEEE) June 1981, pp. 518–524.

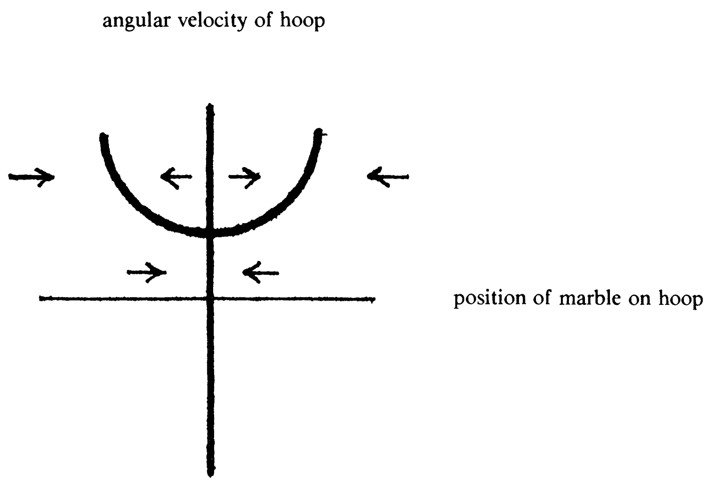
Elementary Stability and Bifurcation Theory. By Gérard Iooss and Daniel D. Joseph. Springer-Verlag, New York, 1980. xv + 286 pp.

STEPHEN SCHECTER
Department of Mathematics, North Carolina State University, Raleigh, NC 27650

Bifurcation is the appearance of new solutions to a problem when a parameter changes. A simple example is what happens (or at least what one can imagine happens) when one places a marble in the groove of an upright hula hoop and spins the hoop about its vertical axis.*

slow fast

If the hoop does not spin too fast, the marble sits at the bottom of the hoop. If the hoop spins faster than a critical angular velocity, the marble creeps up the hoop, to a point where gravity and centrifugal force balance, and stays there. One cannot predict to which side of the hoop the marble will move. Thus we have the "pitchfork" bifurcation diagram:



*This example is taken from J. Marsden and M. McCracken, *The Hopf Bifurcation and its Applications*, Springer-Verlag, 1976, p. 3.

The bottom position remains a solution for all angular velocities of the hoop, i.e., if the marble sits precisely at the bottom of the hoop, there is no force on it. However, when the angular velocity of the hoop passes a critical value, the bottom solution becomes unstable, and two new stable solutions—one of which we actually observe—appear. (Since the force on the marble depends on the marble's velocity relative to the hoop as well as its position and the angular velocity of the hoop, the state space for this problem actually has another dimension. Nevertheless the diagram captures the essential features.)

The study of bifurcations has been invigorated during the past twenty years by the injection of ideas from various areas of mathematics. These ideas include topological degree and related topological invariants; symmetry groups; methods of proof that depend on the center manifold theorem and hard implicit function theorems; the notions of contact equivalence, normal forms, and unfolding from the theory of singularities of maps; and the understanding of complicated qualitative behavior of solutions of differential equations achieved by dynamical systems theory. At the same time the traditional power series methods for computing bifurcating solutions and determining their stability have been refined and applied to new situations. Iooss and Joseph's book is mostly about power series methods.

Let us work out an example. In the hoop-and-marble example we were concerned with bifurcation of equilibrium solutions of an ordinary differential equation defined on a finite dimensional space. It is more usual in Iooss and Joseph's book to be concerned with parabolic partial differential equations. So for our example we shall look for equilibrium solutions of

$$\begin{gathered} u_t = u_{xx} + (1+\lambda)u + u^2 \\ u(0) = u(\pi) = 0. \end{gathered}$$

Here $u(x,t)$ is a function of $x \in [0,\pi]$ and time t; λ is a parameter. We are interested in equilibrium solutions ($u_t = 0$), so we have

$$\begin{gathered} u'' + (1+\lambda)u + u^2 = 0 \\ u(0) = u(\pi) = 0. \end{gathered} \tag{1}$$

Now u is a function of x only. For each fixed λ (1) is a nonlinear boundary value problem which has at least the trivial solution $u = 0$.

We shall first set up this problem as an analyst might, then formally compute a solution using power series, then indicate how the computation can be justified using the analyst's framework.

We shall look for solutions of (1) in the space of functions u on $[0,\pi]$ such that (i) u is differentiable, (ii) u' is absolutely continuous, (iii) the function u'' of which u' is the antiderivative is square integrable, and (iv) $u(0) = u(\pi) = 0$. Analysts denote this space by $H^2 \cap H_0^1$; we shall call it H. Every function in H can of course be expanded in a Fourier sine series; in fact,

$$H = \left\{ u : [0,\pi] \to \mathbb{R} : u(x) = \sum_1^\infty a_n \sin nx \text{ and } \sum_1^\infty n^4 a_n^2 < \infty \right\}.$$

For a norm on H one may take

$$\|u\| = \left[\sum_1^\infty n^4 a_n^2 \right]^{1/2}.$$

Define $F\colon H \times \mathbb{R} \to L^2[0,\pi]$ by $F(u,\lambda) = u'' + (1+\lambda)u + u^2$. We shall think of $L^2[0,\pi]$ as $\{\Sigma_1^\infty a_n \sin nx : \Sigma_1^\infty a_n^2 < \infty\}$. F is a continuous linear mapping, $(u,\lambda) \to u'' + u$, plus a continuous quadratic mapping, $(u,\lambda) \to \lambda u + u^2$. (The continuity of $u \to u^2$ follows from a Sobolev inequality.) So F is an analytic map. We want solutions of $F(u,\lambda) = 0$ besides the trivial solutions $u = 0$.

We shall look for a curve of nontrivial solutions defined near $u = 0, \lambda = 0$, i.e., a mapping $(u(\varepsilon), \lambda(\varepsilon))$: an interval around 0 in $\mathbb{R} \to H \times \mathbb{R}$ such that $u(\varepsilon) = 0$ if and only if $\varepsilon = 0$, $\lambda(0) = 0$, and $(u(\varepsilon), \lambda(\varepsilon))$ satisfies (1) for each ε. If $u(\varepsilon)$ and $\lambda(\varepsilon)$ are analytic functions of ε, we

can write

$$(u(\varepsilon), \lambda(\varepsilon)) = \sum_{1}^{\infty} \frac{\varepsilon^k}{k!}(u^{(k)}, \lambda^{(k)}). \tag{2}$$

Each coefficient $u^{(k)} \in H$ and each $\lambda^{(k)} \in \mathbb{R}$. Of course there are many ways to parameterize a curve, such as by arc length or by projection onto some axis. In our computation we shall assume our curve can be parameterized by the first Fourier coefficient of u, so ε = first Fourier coefficient of $u(\varepsilon)$.

To determine $u^{(k)}$ and $\lambda^{(k)}$, we substitute (2) into (1), group by powers of ε, and set the coefficients equal to 0:

Coefficient of ε : $u^{(1)\prime\prime} + u^{(1)} = 0$. The only solutions in H are $u^{(1)} = C \sin x$. Since ε is to be the first Fourier coefficient of $u(\varepsilon)$, we set $u^{(1)} = \sin x$.

Coefficient of $\frac{\varepsilon^2}{2!}$: We find that

$$u^{(2)\prime\prime} + u^{(2)} = -2\lambda^{(1)}u^{(1)} - 2[u^{(1)}]^2. \tag{3}$$

Remember, we have not yet chosen $\lambda^{(1)}$. This is a good thing, because whether or not (3) can be solved for $u^{(2)}$ depends on what $\lambda^{(1)}$ is. Suppose the right-hand side of (3) is $\Sigma_1^\infty b_n \sin nx$. Writing $u^{(2)} = \Sigma_1^\infty a_n \sin nx$, we find that

$$\sum_{1}^{\infty} (1 - n^2) a_n \sin nx = \sum_{1}^{\infty} b_n \sin nx.$$

If $b_1 \neq 0$, there are no solutions; if $b_1 = 0$, the general solution is

$$u^{(2)} = \sum_{2}^{\infty} \frac{b_n}{1 - n^2} \sin nx + C \sin x. \tag{4}$$

(This is the Fredholm alternative for $u \to u'' + u$ from H to $L^2[0, \pi]$.) Since $u^{(1)} = \sin x$ we may choose $\lambda^{(1)}$ so that the right hand side of (3), when expanded in a Fourier sine series, has $b_1 = 0$. We then take $u^{(2)}$ to be the solution (4) with $C = 0$ (because of our decision about how to parameterize the curve).

In general, when we look at the coefficient of $\varepsilon^k/k!$, we find that

$$u^{(k)\prime\prime} + u^{(k)} = -k\lambda^{(k-1)}u^{(1)} + \text{terms involving } u^{(1)}, \ldots, u^{(k-1)}, \lambda^{(1)}, \ldots, \lambda^{(k-2)} \tag{5}$$
$$\text{(which have already been determined).}$$

We choose $\lambda^{(k-1)}$ so that (5) has solutions, then take $u^{(k)}$ to be the solution whose first Fourier coefficient is 0.

So we have a formal solution (2). To understand it better, we shall look at $F(u, \lambda) = 0$ from a more geometric viewpoint. We can use the implicit function theorem to reduce the problem of solving $F(u, \lambda) = 0$ to just one real equation in two real unknowns, and we can easily determine enough about this equation to see that it has a unique nontrivial analytic solution curve through $u = 0, \lambda = 0$. This is the curve whose Taylor expansion we have found, so our solution is not just formal. Here is the argument:

The derivative of F at (u, λ) is a linear map from $H \times \mathbb{R}$ to $L^2[0, \pi]$. We have $DF(0,0) \cdot (v, \mu) = v'' + v$. We let

K = span of $\sin x$ in H.

$K^\perp$ = subspace of H consisting of functions whose first Fourier coefficient is 0.

R = subspace of $L^2[0,1]$ consisting of functions whose first Fourier coefficient is 0.

$R^\perp$ = span of $\sin x$ in $L^2[0,1]$.

The linear map $v \to v'' + v$ from H to $L^2[0, \pi]$ has kernel K and range R. (If this linear map were an isomorphism, the implicit function theorem would tell us there could be no nontrivial solutions near $u = 0, \lambda = 0$.) Let $P : L^2[0, \pi] \to R$ denote projection onto R with kernel $R^\perp$, so that

$$P\left(\sum_1^\infty b_n \sin nx\right) = \sum_2^\infty b_n \sin nx \text{ and } (I - P)\left(\sum_1^\infty b_n \sin nx\right) = b_1 \sin x.$$

Now $F(u, \lambda) = 0$ if and only if $PF(u, \lambda) = 0$ and $(I - P)F(u, \lambda) = 0$. Since $P \cdot DF(0,0)$ is surjective, the implicit function theorem tells us that near $(0,0)$ the set of (u, λ) such that $PF(u, \lambda) = 0$ is a two-dimensional surface tangent to $K \times \mathbb{R}$. In other words, $PF(u, \lambda) = 0$ if and only if

$$u = (\alpha \sin x + v(\alpha, \lambda), \lambda),$$

where $v(\alpha, \lambda)$ is an analytic mapping of $\mathbb{R}^2$ into $K^\perp$, $v(0,0) = 0$, and $Dv(0,0) = 0$. To solve $F(u, \lambda) = 0$ we need only solve $(I - P)F(u, \lambda) = 0$ where (u, λ) has the above form. This is just one analytic equation in two unknowns. (This reduction is called the Liapunov-Schmidt procedure.)

Writing $(I - P)F(\alpha \sin x + v(\alpha, \lambda), \lambda) = 0$, we find

(6) $$(I - P)(\lambda\alpha \sin x + \alpha^2 \sin^2 x + \text{higher order terms}) = 0.$$

Equation (6) just says that the first Fourier coefficient of the function in parentheses is 0:

(7) $$\frac{2}{\pi}\int_0^\pi (\lambda\alpha \sin x + \alpha^2 \sin^2 x + \cdots)\sin x\, dx = 0$$

$$\lambda\alpha + \frac{8}{3\pi}\alpha^2 + \cdots = 0.$$

An equation in two variables of the form (7) has for its solution set near $(0,0)$ two analytic curves, one tangent to $\alpha = 0$, one tangent to $\lambda + 8\alpha/3\pi = 0$. The first is just the trivial solution, the second is the curve of nontrivial solutions we computed earlier. This justifies our power series computation and shows, moreover, that there are no other solutions near $u = 0, \lambda = 0$.

Most of Iooss and Joseph's book is careful computation of various kinds of bifurcating solutions and their stability by power series methods. The variety of problems that can be handled by this technique is impressive. I want to mention in particular the treatment of subharmonic bifurcation, which is the bifurcation, from a periodic solution of a periodically forced differential equation, of periodic solutions with longer period. The justification of power series methods, which may require difficult analysis, is only sketched. A mathematician—even one familiar with the "modern" approaches to bifurcation theory—can learn a lot from Iooss and Joseph's book, but he will have to fill in some theory. The book's natural audience is applied mathematicians, engineers and scientists, at the beginning graduate level on up, who are motivated to learn the methods by bifurcation problems they have encountered in their work. Springer-Verlag has included Iooss and Joseph's book in its undergraduate text series, but the book is probably too demanding for an undergraduate course.

MISCELLANEA

111. You state this result and then try to mortify me by saying that it is easy to prove. Since I can't succeed in doing so, I appeal to your good nature to help me out of the difficulty.

—Hermite to Stieltjes, 26 April 1892.

The linear map $v \to v'' + v$ from H to $L^2[0, \pi]$ has kernel K and range R. (If this linear map were an isomorphism, the implicit function theorem would tell us there could be no nontrivial solutions near $u = 0, \lambda = 0$.) Let $P : L^2[0, \pi] \to R$ denote projection onto R with kernel $R^\perp$, so that

$$P\left(\sum_1^\infty b_n \sin nx\right) = \sum_2^\infty b_n \sin nx \text{ and } (I - P)\left(\sum_1^\infty b_n \sin nx\right) = b_1 \sin x.$$

Now $F(u, \lambda) = 0$ if and only if $PF(u, \lambda) = 0$ and $(I - P)F(u, \lambda) = 0$. Since $P \cdot DF(0,0)$ is surjective, the implicit function theorem tells us that near $(0,0)$ the set of (u, λ) such that $PF(u, \lambda) = 0$ is a two-dimensional surface tangent to $K \times \mathbb{R}$. In other words, $PF(u, \lambda) = 0$ if and only if

$$u = (\alpha \sin x + v(\alpha, \lambda), \lambda),$$

where $v(\alpha, \lambda)$ is an analytic mapping of $\mathbb{R}^2$ into $K^\perp$, $v(0,0) = 0$, and $Dv(0,0) = 0$. To solve $F(u, \lambda) = 0$ we need only solve $(I - P)F(u, \lambda) = 0$ where (u, λ) has the above form. This is just one analytic equation in two unknowns. (This reduction is called the Liapunov-Schmidt procedure.)

Writing $(I - P)F(\alpha \sin x + v(\alpha, \lambda), \lambda) = 0$, we find

$$(I - P)(\lambda\alpha \sin x + \alpha^2 \sin^2 x + \text{higher order terms}) = 0. \tag{6}$$

Equation (6) just says that the first Fourier coefficient of the function in parentheses is 0:

$$\frac{2}{\pi}\int_0^\pi (\lambda\alpha \sin x + \alpha^2 \sin^2 x + \cdots)\sin x \, dx = 0 \tag{7}$$

$$\lambda\alpha + \frac{8}{3\pi}\alpha^2 + \cdots = 0.$$

An equation in two variables of the form (7) has for its solution set near $(0,0)$ two analytic curves, one tangent to $\alpha = 0$, one tangent to $\lambda + 8\alpha/3\pi = 0$. The first is just the trivial solution, the second is the curve of nontrivial solutions we computed earlier. This justifies our power series computation and shows, moreover, that there are no other solutions near $u = 0, \lambda = 0$.

Most of Iooss and Joseph's book is careful computation of various kinds of bifurcating solutions and their stability by power series methods. The variety of problems that can be handled by this technique is impressive. I want to mention in particular the treatment of subharmonic bifurcation, which is the bifurcation, from a periodic solution of a periodically forced differential equation, of periodic solutions with longer period. The justification of power series methods, which may require difficult analysis, is only sketched. A mathematician—even one familiar with the "modern" approaches to bifurcation theory—can learn a lot from Iooss and Joseph's book, but he will have to fill in some theory. The book's natural audience is applied mathematicians, engineers and scientists, at the beginning graduate level on up, who are motivated to learn the methods by bifurcation problems they have encountered in their work. Springer-Verlag has included Iooss and Joseph's book in its undergraduate text series, but the book is probably too demanding for an undergraduate course.

MISCELLANEA

111. You state this result and then try to mortify me by saying that it is easy to prove. Since I can't succeed in doing so, I appeal to your good nature to help me out of the difficulty.

—Hermite to Stieltjes, 26 April 1892.

LETTERS TO THE EDITOR

Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

When Professor Mark Kac wrote the review (this *Monthly,* 89 (1982) 713–714) of my double-biographical essay, *John von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death,* he evidently based it on some misinformation. The purpose of this note is the set the record straight.

Kac is under the mistaken impression that I never met Mr. Nicholas Vonneuman, John von Neumann's brother. In fact Mr. Vonneuman and I spent two hours together over lunch well over a decade ago, on May 18, 1971, in Philadelphia. The purpose of our lunch meeting was primarily to get acquainted, since already at that time I was planning to write about John von Neumann.

—S. J. Heims
Department of Physics
University of Massachusetts, Boston.

Professor Kac comments as follows:

When I learned from a third person that Mr. Vonneumann claimed never to have met Mr. Heims, I asked Mr. Vonneumann for permission to include this information in the review and I sent him a copy of the review. Though no formal permission has been given (on the grounds that it was unnecessary), there has been no objection on the part of Mr. Vonneumann to its inclusion.

—Mark Kac

Editor:

I have been trying to document the reason for there not being a Nobel Prize in Mathematics. Is this actually due to a feud between Nobel and Mittag-Leffler?

I should be interested in hearing from someone who has a reference on this point.

—Amy C. King
Department of Mathematics
Eastern Kentucky University
Richmond, Kentucky 40475

If anyone can supply some documentation that is not just iterated gossip, probably most readers of the MONTHLY would like to know about it. It might be quite appropriate to publish an answer to Professor King's question as another Letter to the Editor.

—Editor

AMS SHORT COURSE LECTURE NOTES
A Subseries in Proceedings of Symposia in Applied Mathematics

The AMS Short Course Series was initiated in 1973 in response to what was then an urgent employment problem for recent mathematics Ph.D's. It has turned out to have a wider impact. The applied mathematical sciences are currently experiencing real growth and indications are that this trend will continue. The Short Course series addresses the problem of keeping mathematicians in touch with new developments for both intellectual and pragmatic reasons.

Computed Tomography
Lawrence A. Shepp, Editor

Lawrence A. Shepp, *Scope of pure and applied tomography*
Kennan T. Smith, *Reconstruction formulas in computed tomography*
Oleh J. Tretiak, *Attenuated and exponential Radon transforms*
Allan M. Cormack, *Computed tomography: Some history and recent developments*
F. Alberto Grünbaum, *The limited angle reconstruction problem*
Sigurdur Helgason, *Ranges of Radon transforms*
Johann Radon, *Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten* (reprinted from *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig* 69 (1917), 262–277).

Volume 27, x + 86 pages (hard cover and soft cover)
Soft cover prices: List $14, institutional member $11, individual member $7
Hard cover prices: List $20, institutional member $15, individual member $10
ISBN 0-8218-0033-7; LC 82-18508
Publication date: January 1983
To order, please specify PSAPM/27MA (hard cover)
PSAPMS/27MA (soft cover)

The Mathematics of Networks
Stefan A. Burr, Editor

Frank Boesch, *Introduction to basic network problems*
Frances Yao, *Maximum flows in networks*
R. M. Karp, *The computational complexity of network problems*
Shen Lin, *Effective use of heuristic algorithms in network design*
Daniel J. Kleitman, *Some practical network problems*
Nicholas Pippenger, *Telephone switching networks*
Stefan A. Burr, *Concluding remarks*

Volume 26, viii + 142 pages (hard cover and soft cover)
Soft cover prices: List $16, institutional member $16, individual member $8
Hard cover prices: List $22, institutional member $12, individual member $11
ISBN 0-8218-0031-0; LC 82-18469
Publication date: December 1982
To order, please specify PSAPM/26 MA (hard cover)
PSAPMS/26 MA (soft cover)

EARLIER VOLUMES

MATHEMATICAL ASPECTS OF PRODUCTION AND DISTRIBUTION OF ENERGY
edited by *Peter D. Lax*

Volume 21, 137 pages (soft cover)
List price $17, institutional member $13, individual member $9
ISBN 0-8218-0121-X; LC 77-7174
Publication date: 1977, reprinted with corrections 1979
To order, please specify PSAPM/21MA

NUMERICAL ANALYSIS
edited by *Gene H. Golub* and *Joseph Oliger*

Volume 22, viii + 135 pages (soft cover)
List price $12, institutional member $9, individual member $6
ISBN 0-8218-0122-8; LC 78-11096,
Publication date: 1978
To order, please specify PSAPM/22MA

MODERN STATISTICS: METHODS AND APPLICATIONS
edited by *Robert V. Hogg*

Volume 23, vi + 110 pages (soft cover)
List price $14, institutional member $11, individual member $7
ISBN 0-8218-0023-X; LC 80-16093
Publication date: July 1980
To order, please specify PSAPM/23MA

GAME THEORY AND ITS APPLICATIONS
edited by *William F. Lucas*

Volume 24, viii + 128 pages (soft cover)
List price $13, institutional member $10, individual member $7
ISBN 0-8218-0025-6; LC 81-12914
Publication date: October 1981
To order, please specify PSAPM/24MA

OPERATIONS RESEARCH: MATHEMATICS AND MODELS
edited by *Saul I. Gass*

Volume 25, viii + 198 pages (soft cover)
List price $11, institutional member $8, individual member $6
ISBN 0-8218-0029-9; LC 81-10849
Publication date: September 1981
To order, please specify PSAPM/25MA

Prepayment is required for all AMS publications. Order from AMS, P.O. Box 1571, Annex Station, Providence, RI 02901, or call toll free 800-556-7774 to charge with Visa or MasterCard.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 8 October 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

THE QUARTERLY REPORTS OF S. RAMANUJAN

BRUCE C. BERNDT*

Department of Mathematics, University of Illinois, Urbana, IL 61801

Introduction. From about the age of 15 or 16 until his untimely death at the age of 32, Ramanujan devoted all of his energy to the pursuit of mathematics. His research was accomplished in isolation and obscurity until, at the age of 22, he obtained a meeting with V. R. Aiyar, the founder of the Indian Mathematical Society. From that moment in 1910, word of Ramanujan's mathematical genius slowly began to spread among mathematicians in southeast India. Several people, including R. Ramachandra Rao, P. V. Seshu Aiyar, S. N. Aiyar, Sir Francis Spring, and Sir Gilbert Walker, took a kindly interest in Ramanujan through financial support, employment, and encouragement. In particular, on February 26, 1913, the English astronomer Walker sent a letter to the registrar of the University of Madras, Francis Dewsbury, with the emphatic recommendation, "The University would be justified in enabling S. Ramanujan for a few years *at least* to spend the whole of his time on mathematics, without any anxiety as to his livelihood." The Board of Studies at the University of Madras agreed to this request, and its chairman, Professor B. Hanumantha Rao, wrote a letter to the Vice-Chancellor on March 25, 1913, with an exhortation that Ramanujan be awarded a scholarship of 75 rupees per month. Again, the decision was swift, and Ramanujan was granted a scholarship commencing on May 1, 1913.

A stipulation in the scholarship required Ramanujan to submit quarterly reports detailing his research to the Board of Studies in Mathematics. Ramanujan wrote three such quarterly reports dated 5th August 1913, 7th November 1913, and 9th March 1914 before he departed for England on March 17, 1914. Possibly these reports still remain at the University of Madras, but they evidently have been either lost or misplaced. Fortunately, in 1925, T. A. Satagopan made a handwritten copy of the reports on 51 foolscap pages. This copy was sent to G. H. Hardy and is now at the library of Trinity College, Cambridge. Also on file at Trinity College is a second copy of the reports made by G. N. Watson, who, along with B. M. Wilson, attempted to edit Ramanujan's notebooks [22] in the 1930's. Although the reports have never been published, Hardy used material from the reports as the basis for Chapter 11 of his book [14] on Ramanujan's work.

Besides the quarterly reports, other manuscripts of Ramanujan remain unpublished. Two quite different and fascinating descriptions of some of these papers have recently been given by K. G. Ramanathan [19] and R. A. Rankin [23]. An interesting account of Ramanujan's "lost" notebook has been written for this MONTHLY by G. E. Andrews [1], who is in the process of attempting to prove all of the formulas in this manuscript. An unedited facsimile edition of Ramanujan's notebooks [22] has been published. For general descriptions of the notebooks, see papers of Watson [25], and the author [3] who currently is editing chapters of the notebooks.

The purpose of this paper is to describe the most significant results found in the quarterly reports and to place them in an historical perspective. A complete description of the reports is being published elsewhere [4]. In contrast to his notebooks which contain very few proofs, or even sketches, the quarterly reports offer several fairly detailed proofs. However, many of these proofs, especially those for the principal theorems, are formal and not rigorous. Nonetheless, Ramanujan's proofs are enormously interesting because they provide insight into how Ramanujan reasoned,

---

The author received his A.B. degree from Albion College in 1961 and his Ph.D. from the University of Wisconsin in 1966 under the direction of J. R. Smart. After one year at the University of Glasgow, he came to the University of Illinois at Urbana-Champaign where he has remained, except for one year at the Institute for Advanced Study. His major research interests are analytic number theory and classical analysis. Since 1977 he has spent almost all of his research efforts on editing Ramanujan's notebooks, a task which will require at least five more years to complete. In 1979 he received an Allendoerfer award from the MAA for his paper, Ramanujan's Notebooks.

*Research partially supported by the Vaughn Foundation.

and so we shall describe some of Ramanujan's arguments. We shall also indicate, frequently with references to the literature, how Ramanujan's theorems can be put on firm foundations.

The first two reports and a portion of the third are concerned with certain integral formulas which, in a sense, are interpolation formulas and which are connected with the theory of integral transforms. In discussing one of these formulas in an article for the MONTHLY in 1937, Hardy [**13**, p. 150], [**14**, Chapter 1] remarks, "There is one particularly interesting formula... of which he was especially fond and made continual use." In Chapter 11 of [**14**], Hardy further observes, in connection with the aforementioned formulas, that Ramanujan "had not 'really' proved any of the formulae which I have quoted. It was impossible that he should have done so because the 'natural' conditions involve ideas of which he knew nothing in 1914, and which he had hardly absorbed before his death." The natural conditions to which Hardy refers are from the theory of functions of a complex variable. Indeed, as we shall momentarily see, the proofs given in the quarterly reports are not rigorous. After arriving in England, Ramanujan evidently learned, probably from Hardy, that his proofs were not sound. For in a paper published in 1915, after offering some beautiful integral evaluations, Ramanujan [**20**], [**21**, p. 57] remarks that "My own proofs of the above results make use of a general formula, the truth of which depends on conditions which I have not yet investigated completely. A direct proof depending on Cauchy's theorem will be found in Mr. Hardy's note which follows this paper."

Some of the principal integral formulas and their applications in the quarterly reports can be found in the notebooks [**22**], primarily in Chapters 3 and 4, which have now been thoroughly examined and edited [**5**], [**6**]. As remarked above, some of the reports' contents have been discussed by Hardy [**14**]. However, the quarterly reports contain many additional results not covered in the aforementioned sources. In particular, a beautiful and very useful generalization of Frullani's integral theorem is found in the reports. This hitherto unobserved theorem provides a powerful tool for evaluating many integrals and deserves more attention.

In addition to material on integral interpolation formulas and transforms, the third quarterly report contains results on orders of infinity, fractional composition of functions, and fractional differentiation. Each of these topics is treated only briefly. It might be recalled that in his first letter to Hardy [**21**, p. xxiii], dated 16th January 1913, Ramanujan mentions that he has been reading Hardy's tract, Orders of Infinity [**12**].

The author is grateful to the Master and Fellows at Trinity College, Cambridge, for a copy of the quarterly reports and permission to publish an analysis of them.

1. The First Quarterly Report. In contrast to the second and third reports, the first report commences with a letter of introduction which we completely reproduce below.

> From S. Ramanujan, Scholarship holder in mathematics.
> To the Board of Studies in Mathematics.
> Through the Registrar, University of Madras.
> Gentlemen,
>
> With reference to para. 2 of the University Registrar's letter no. 1631 dated the 9th April 1913, I beg to submit herewith my quarterly Progress Report for the quarter ended the 31st July, 1913.
>
> The Progress Report is merely the exposition of a new theorem I have discovered in Integral Calculus. At present there are many definite integrals the values of which we know to be finite but still not possible of evaluation by the present known methods. This theorem will be an instrument by which at least some of the definite integrals whose values are at present not known can be evaluated. For instance, the integral treated in Ex(v) *note* Art. 5 in the paper, Mr. G. H. Hardy, M.A., F.R.S. of Trinity College, Cambridge, considers to be "new and interesting." Similarly the integral connected with the Besselian function of the nth order which at present requires many complicated manipulations to evaluate can be readily inferred from the theorem given in the paper. I have also utilised this theorem in definite integrals for the expansion of functions which can now be ordinarily done by Lagrange's, Burmann's, or Abel's theorems. For instance, the expansions marked as examples nos. (3) and (4), Art. 6, in the second part of the paper.

The investigations I have made on the basis of this theorem are not all contained in the attached paper. There is ample scope for new and interesting results out of this theorem. This paper may be considered the first installment of the results I have got out of the theorem. Other new results based on the theorem I shall communicate in my later reports.

I beg to submit this, my maiden attempt, and I humbly request that the Members of the Board will make allowance for any defect which they may notice to my want of usual training which is now undergone by college students and view sympathetically my humble effort in the attached paper.

I beg to remain,
Gentlemen
Your obedient servant
S. Ramanujan

We now describe the primary theorem of the quarterly reports. We emphasize that the hypotheses below are those supplied by Ramanujan and need to be strengthened to insure the stated conclusions.

THEOREM 1 (Ramanujan's Master Theorem). *Suppose that, in some neighborhood of* $x = 0$,

$$F(x) = \sum_{k=0}^{\infty} \frac{\phi(k)(-x)^k}{k!}.$$

Then

$$I \equiv \int_0^{\infty} x^{n-1} F(x)\, dx = \Gamma(n)\phi(-n), \tag{1.1}$$

where Γ *denotes the gamma function. Conversely, if the value of* I *is given by* (1.1), *then the Maclaurin coefficients of* $F(x)$ *can be determined.*

Ramanujan's Master Theorem needs some explanation. Generally, n is not a nonpositive integer. Thus, Ramanujan is assuming that there exists a continuous extension of ϕ, defined initially on the set of nonnegative integers. Ramanujan provides lengthy arguments in order to convince the reader that there always exists a unique, "natural," continuous extension of ϕ, but, as to be expected, his pleas are unconvincing. Ramanujan requires just four simple hypotheses for his theorem: (1) $F(x)$ can be expanded in a Maclaurin series, (2) $F(x)$ is continuous on $(0, \infty)$, (3) $n > 0$, and (4) $x^n F(x)$ tends to 0 as x tends to ∞. Ramanujan remarks that the fourth condition can be relaxed if F is a circular function.

Not surprisingly, more stringent hypotheses need to be put on F and ϕ for such a theorem to exist. We shall first give Ramanujan's "proof" and then we shall state Hardy's rigorous reformulation of Ramanujan's theorem.

Ramanujan's Proof. Recall Euler's integral representation of the gamma function

$$\int_0^{\infty} e^{-mx} x^{n-1}\, dx = m^{-n}\Gamma(n),$$

where $m, n > 0$. Let $m = r^k$ with $r > 0$, multiply both sides by $f^{(k)}(a)h^k/k!$, where f shall be specified later, and sum on k, $0 \leqslant k < \infty$, to obtain

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)h^k}{k!} \int_0^{\infty} e^{-r^k x} x^{n-1}\, dx = \Gamma(n) \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(hr^{-n})^k}{k!}.$$

Next, expand $\exp(-r^k x)$, $0 \leqslant k < \infty$, in its Maclaurin series, invert the order of summation and integration, invert the order of summation, and apply Taylor's theorem to deduce that

$$\int_0^{\infty} x^{n-1} \sum_{j=0}^{\infty} \frac{f(hr^j + a)(-x)^j}{j!}\, dx = \Gamma(n) f(hr^{-n} + a). \tag{1.2}$$

Now define $f(hr^m + a) = \phi(m)$, where m is real, and a, h, and r are regarded as constants. Then (1.2) may be rewritten in the form

$$\int_0^\infty x^{n-1} \sum_{j=0}^\infty \frac{\phi(j)(-x)^j}{j!} dx = \Gamma(n)\phi(-n),$$

which completes Ramanujan's proof.

Ramanujan was evidently quite fond of the clever, original technique described above, and he employed it in many contexts. See [**6**] for several additional illustrations.

In preparation for stating Hardy's version of Ramanujan's theorem, we need to introduce some notation. Let $s = \sigma + it$ with σ and t both real. Let $H(\delta) = \{s : \sigma \geqslant -\delta\}$, where $0 < \delta < 1$. Suppose that $\psi(s)$ is analytic on $H(\delta)$ and that there exist constants C, P, and A with $A < \pi$ such that

$$|\psi(s)| \leqslant Ce^{P\sigma + A|t|}, \tag{1.3}$$

for all $s \in H(\delta)$. For $x > 0$ and $0 < c < \delta$, define

$$\Psi(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\pi}{\sin(\pi s)}\psi(-s)x^{-s}\, ds.$$

If $0 < x < e^{-P}$, an application of the residue theorem yields [**14**, p. 189]

$$\Psi(x) = \sum_{k=0}^\infty \psi(k)(-x)^k.$$

THEOREM (Hardy [**14**, pp. 189, 190]). *Let ψ and Ψ satisfy the conditions set forth in the preceding paragraph. Suppose that* $0 < \sigma < \delta$. *Then*

$$\int_0^\infty \Psi(x)x^{s-1}\, dx = \frac{\pi}{\sin(\pi s)}\psi(-s). \tag{1.4}$$

Formula (1.4) yields (1.1) upon replacing $\psi(s)$ by $\phi(s)/\Gamma(s+1)$. In Hardy's book [**14**], formulas (1.1) and (1.4) are (B) and (A), respectively, on p. 186.

Ramanujan devotes the remainder of the first quarterly report, all of the second report, and a large portion of the third to giving examples and applications of his Master Theorem. We shall present some of the more interesting examples in the sequel. In some instances, Hardy's theorem can be invoked to provide a legitimate proof. In other examples, Hardy's theorem does not apply. However, in every case a rigorous proof can be given, and so we shall often cite the literature where proofs may be found.

EXAMPLE 1.1. Let $m, n > 0$ and set $x = y/(1+y)$ to obtain

$$\int_0^1 x^{m-1}(1-x)^{n-1}\, dx = \int_0^\infty y^{m-1}(1+y)^{-m-n} dy.$$

From the binomial series,

$$(1+y)^{-r} = \sum_{k=0}^\infty \frac{\Gamma(k+r)}{\Gamma(r)k!}(-y)^k, \quad |y| < 1, \tag{1.5}$$

we find that $\phi(s) = \Gamma(s+m+n)/\Gamma(m+n)$. By Stirling's formula, the hypothesis (1.3) of Hardy's theorem is readily verified. Hence, Ramanujan's Master theorem yields the following well-known representation of the beta-function,

$$B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}\, dx = \Gamma(m)\phi(-m) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}.$$

EXAMPLE 1.2. For $0 < q < 1$, define

$$(a;q)_\infty = \prod_{k=0}^{\infty}(1-aq^k)$$

and

$$I = \int_0^\infty \frac{t^{n-1}\,dt}{(-t;q)_\infty}, \quad n > 0,$$

which is actually the particular case $a = 0$ of the more general integral

$$f(a) = \int_0^\infty \frac{t^{n-1}(-at;q)_\infty\,dt}{(-t;q)_\infty}, \quad |a| < q^n.$$

This is one of those integrals in **[20]**, **[21**, pp. 53–58] to which we referred in the Introduction. Also, $f(a)$ is that integral mentioned by Ramanujan in his prefatory comments. For a thorough discussion of $f(a)$, see a recent article in this MONTHLY by R. A. Askey **[2]**.

We now describe Ramanujan's proof, which is also briefly sketched by Hardy **[14**, p. 194]. Writing

$$\frac{1}{(-t;q)_\infty} = \sum_{k=0}^{\infty}\psi(k)(-t)^k, \quad |t| < 1,$$

and using the fact that $(1+t)(-qt;q)_\infty = (-t;q)_\infty$, we easily derive the recursion formula

$$\psi(k) = \frac{\psi(k-1)}{1-q^k}, \quad k \geqslant 1.$$

Since $\psi(0) = 1$, an inductive argument shows that

$$\psi(k) = \frac{(q^{k+1};q)_\infty}{(q;q)_\infty}, \quad k \geqslant 0,$$

which satisfies the hypotheses of Hardy's theorem. Hence, by Ramanujan's Master theorem, or Hardy's theorem,

$$I = \frac{\pi}{\sin(\pi n)}\psi(-n) = \frac{\pi(q^{1-n};q)_\infty}{\sin(\pi n)(q;q)_\infty}.$$

EXAMPLE 1.3. Let $a \geqslant 0$ and let x be the unique positive solution to the equation $\operatorname{Log} x = -ax$. For each positive number n, we want to expand x^n in ascending powers of a. Letting $0 < p < n$, $a = -(\operatorname{Log} x)/x$, and then $x = e^{-y}$, Ramanujan finds that

$$\begin{aligned}\int_0^\infty a^{p-1}x^n\,da &= \int_0^1\left(-\frac{\operatorname{Log} x}{x}\right)^{p-1} x^n\frac{1-\operatorname{Log} x}{x^2}\,dx\\ &= \int_0^\infty y^{p-1}(1+y)e^{-y(n-p)}\,dy\\ &= \frac{n\Gamma(p)}{(n-p)^{p+1}}.\end{aligned}$$

Thus, in the notation of the Master Theorem, $\phi(p) = n(n+p)^{p-1}$. Therefore, Ramanujan concludes that

$$x^n = n\sum_{k=0}^{\infty}\frac{(n+k)^{k-1}(-a)^k}{k!}. \tag{1.6}$$

Using Stirling's formula, one can show that the infinite series in (1.6) converges for $0 \leqslant a \leqslant 1/e$. In fact, (1.6) is valid for every real number n and $|a| \leqslant 1/e$. The expansion (1.6) can be found in

Chapter 3 of Ramanujan's second notebook [**22**, p. 32], and a rigorous proof has been given in [**5**, Entry 13].

EXAMPLE 1.4. Consider the trinomial equation

$$aqx^p + x^q = 1, \tag{1.7}$$

where $a > 0$ and $0 < q < p$. Using a procedure similar to that in Example 1.3, Ramanujan derives the power series expansion

$$x^n = \frac{n}{q}\sum_{k=0}^{\infty}\frac{\Gamma\left(\frac{n+pk}{q}\right)(-qa)^k}{\Gamma\left(\frac{n+pk}{q}-k+1\right)k!}, \tag{1.8}$$

where $n > 0$ and x is a particular root of (1.7). By Stirling's formula, this expansion converges when

$$|a| \leqslant p^{-p/q}(p-q)^{(p-q)/q}.$$

Equality (1.8) can also be found in Chapter 3 of Ramanujan's second notebook [**22**, vol. 2, p. 33], and a legitimate proof has been given in [**5**, Entry 14].

Formulas (1.6) and (1.8) have a very long history. The latter was first discovered in 1758 by Lambert [**16**], while the former was initially found by Euler [**9**] in 1779. In 1770, Lagrange [**15**] discovered his famous "Lagrange inversion formula" [**18**, p. 145], sometimes called the Lagrange-Bürmann theorem, and as an application derived (1.8). Recall that in the introduction to the first report, Ramanujan makes a reference to theorems of Lagrange and Bürmann. Ramanujan evidently learned about these theorems from Carr's Synopsis [**7**, pp. 278–282], which was the most influential book in Ramanujan's development. Pólya and Szegő [**18**, p. 146] have offered (1.6) and (1.8) as exercises illustrating the Lagrange inversion formula. The derivations of (1.6) and (1.8) in [**5**] are developed *ab initio*, however. For additional references to these formulas, consult [**5**].

2. The Second Quarterly Report. The first part of the second report is devoted to several additional illustrations of the Master Theorem.

EXAMPLE 2.1. We shall suppress the details in Ramanujan's derivation of the very unusual expansion

$$e^{ax} = 1 + \frac{ae^{-bx}\sin(cx)}{c} + \sum_{k=2}^{\infty}\frac{d_k}{k!}\left(\frac{e^{-bx}\sin(cx)}{c}\right)^k, \tag{2.1}$$

where, for $k \geqslant 2$,

$$d_k = \begin{cases} a(a+kb)\prod_{j=1}^{n-1}\{(a+kb)^2+(2jc)^2\}, & \text{if } k = 2n \text{ is even},\\ a\prod_{j=1}^{n-1}\{(a+kb)^2+((2j-1)c)^2\}, & \text{if } k = 2n-1 \text{ is odd}.\end{cases}$$

Here a, b, and c are real with $b, c \geqslant 0$, and $0 \leqslant x \leqslant (1/c)\tan^{-1}(c/b)$. Ramanujan discusses several associated expansions and consequences of (2.1). We mention just one special instance; letting $c = 0$ in (2.1), we find that, for $0 \leqslant x \leqslant 1/b$,

$$e^{ax} = a\sum_{k=0}^{\infty}\frac{(a+kb)^{k-1}x^k e^{-kbx}}{k!}.$$

Ramanujan evidently first learned of this well-known formula from Carr's book [**7**, p. 282]. This

expansion can be used to establish a more general theorem due to Abel [7, p. 282]. This is the theorem to which Ramanujan refers in his introductory letter to the reports.

We now come to perhaps the most significant feature of the quarterly reports, Ramanujan's generalization of Frullani's integral theorem. Suppose that, in some neighborhood of $x = 0$,

$$f(x) - f(\infty) = \sum_{k=0}^{\infty} \frac{u(k)(-x)^k}{k!} \text{ and } g(x) - g(\infty) = \sum_{k=0}^{\infty} \frac{v(k)(-x)^k}{k!},$$

where

$$f(\infty) = \lim_{x \to \infty} f(x) \text{ and } g(\infty) = \lim_{x \to \infty} g(x),$$

which we assume exist.

THEOREM 2. *In the notation above, let f and g be continuous functions on $[0, \infty)$. Assume that $u(s)/\Gamma(s+1)$ and $v(s)/\Gamma(s+1)$ satisfy the hypotheses of Hardy's theorem. Furthermore, assume that $f(0) = g(0)$ and $f(\infty) = g(\infty)$. Then if $a, b > 0$,*

$$\lim_{n \to 0+} I_n \equiv \lim_{n \to 0+} \int_0^{\infty} x^{n-1}\{f(ax) - g(bx)\} dx \tag{2.2}$$
$$= \{f(0) - f(\infty)\}\left\{\operatorname{Log}\frac{b}{a} + \frac{d}{ds}\left(\operatorname{Log}\frac{v(s)}{u(s)}\right)_{s=0}\right\}.$$

Ramanujan tacitly assumes that the limit can be taken under the integral sign. If $f(x) = g(x)$, Theorem 2 reduces to

$$\int_0^{\infty} \frac{f(ax) - f(bx)}{x} dx = \{f(0) - f(\infty)\}\operatorname{Log}(b/a), \tag{2.3}$$

since in this instance the limit on n can always be taken under the integral sign. Formula (2.3) is known as Frullani's theorem and holds for any continuous function f such that $f(\infty)$ exists. If $f(\infty)$ does not exist, but $f(x)/x$ is integrable over $[c, \infty)$ for some $c > 0$, then (2.3) still holds, but with $f(\infty)$ replaced by 0. According to the reports, Ramanujan likely learned of Frullani's theorem from Williamson's book [26] on integral calculus.

Ramanujan's Proof. Applying the Master Theorem with $0 < n < 1$, we find that

$$\begin{aligned} I_n &= \int_0^{\infty} x^{n-1}(\{f(ax) - f(\infty)\} - \{g(bx) - g(\infty)\})\, dx \\ &= \Gamma(n)\{a^{-n}u(-n) - b^{-n}v(-n)\} \\ &= \Gamma(n+1)\left\{\frac{a^{-n}u(-n) - b^{-n}v(-n)}{n}\right\}. \end{aligned}$$

Letting n tend to 0, we deduce that

$$\begin{aligned} \lim_{n \to 0} I_n &= \lim_{n \to 0}\left\{\frac{b^n v(n) - a^n u(n)}{n}\right\} \\ &= \lim_{n \to 0}\{b^n v(n)\operatorname{Log} b + b^n v'(n) - a^n u(n)\operatorname{Log} a - a^n u'(n)\} \\ &= \{f(0) - f(\infty)\}\operatorname{Log}(b/a) + v'(0) - u'(0) \\ &= \{f(0) - f(\infty)\}\left\{\operatorname{Log}(b/a) + \frac{d}{ds}\left(\operatorname{Log}\frac{v(s)}{u(s)}\right)_{s=0}\right\}, \end{aligned} \tag{2.4}$$

where we have used the fact that $u(0) = v(0) = f(0) - f(\infty)$.

We record just one of several examples given by Ramanujan.

EXAMPLE 2.2. Recall again the binomial series (1.5). Using the standard notation $\psi(x) = \Gamma'(x)/\Gamma(x)$, we find from Theorem 2 that, for $a, b, p, q > 0$,

$$(2.5)\quad \int_0^\infty \frac{(1+ax)^{-p} - (1+bx)^{-q}}{x}\,dx = \operatorname{Log}(b/a) + \frac{d}{ds}\operatorname{Log}\left(\frac{\Gamma(q+s)\Gamma(p)}{\Gamma(p+s)\Gamma(q)}\right)_{s=0}$$
$$= \operatorname{Log}(b/a) + \psi(q) - \psi(p)$$
$$= \operatorname{Log}(b/a) + \sum_{k=0}^\infty \left\{\frac{1}{p+k} - \frac{1}{q+k}\right\},$$

since [10, p. 943]

$$\psi(x) = -\gamma - \sum_{k=0}^\infty \left\{\frac{1}{k+x} - \frac{1}{k+1}\right\},$$

where γ denotes Euler's constant. Ramanujan lists four special cases of (2.5), one of which is

$$\int_0^\infty \frac{(1+ax)^{-1/4} - (1+bx)^{-3/4}}{x}\,dx = \operatorname{Log}(b/a) + \pi,$$

where Leibniz's series for $\pi/4$ was employed.

James Hafner has kindly shown us that Theorem 2 can be significantly strengthened. Hafner's improvement allows Ramanujan's formula (2.2) to be applied to a wider variety of functions (and also for the limit in (2.2) to be taken under the integral sign). We now present Hafner's argument.

By Frullani's Theorem (2.3),

$$\int_0^\infty \frac{f(ax) - g(bx)}{x}\,dx = \int_0^\infty \frac{\{f(ax) - f(bx)\} + \{f(bx) - g(bx)\}}{x}\,dx$$
$$= \{f(0) - f(\infty)\}\operatorname{Log}(b/a) + \int_0^\infty \frac{f(x) - g(x)}{x}\,dx.$$

By (2.4), it then remains to show that

$$\int_0^\infty \frac{f(x) - g(x)}{x}\,dx = -\{u'(0) - v'(0)\}.$$

Replacing $f(x) - g(x)$ by $f(x)$, it now suffices to prove the following lemma. (A similar result was established by Carlson [8] under stronger hypotheses.)

LEMMA. *Suppose that f is analytic in a neighborhood of the nonnegative real axis. Put*

$$f(z) = \sum_{k=0}^\infty \frac{u(k)(-z)^k}{k!},$$

for $|z|$ sufficiently small. Assume that, for some positive number δ,

$$\int^\infty f(x)x^{-\alpha}\,dx$$

converges uniformly for $1 - \delta < \operatorname{Re}(\alpha) < 2$. Suppose furthermore that $f(0) = 0$. Then $u(s)$ can be extended in a neighborhood of $s = 0$ so that u is differentiable at $s = 0$ and so that

$$(2.6)\qquad u'(0) = -\int_0^\infty \frac{f(x)}{x}\,dx.$$

Proof. For $\varepsilon > 0$ sufficiently small, we shall, in fact, define u by

$$(2.7)\qquad u(s) = \frac{\Gamma(s+1)e^{\pi i s}}{2\pi i}\oint_{|z|=\varepsilon} f(z)z^{-s-1}\,dz$$

and then show that u has the desired properties. (We shall assume that principal branches are always taken.) First, observe that if s is a nonnegative integer, (2.7) is valid by Cauchy's integral formula for derivatives. Since f is analytic in a neighborhood of the positive real axis, by Cauchy's theorem,

$$\int_{\varepsilon e^{2\pi i}}^{Re^{2\pi i}} f(z) z^{-s-1}\, dz - \int_{\varepsilon}^{R} f(z) z^{-s-1}\, dz = 0, \quad 0 < \varepsilon < R < \infty.$$

It follows from (2.7) that

$$u(s) = \frac{\Gamma(s+1)e^{\pi i s}}{2\pi i}\left\{\oint_{|z|=\varepsilon} f(z) z^{-s-1}\, dz + (e^{-2\pi i s} - 1)\int_{\varepsilon}^{R} f(x) x^{-s-1}\, dx\right\}.$$

Assuming that $-\delta < \mathrm{Re}(s) < 1$ and using our hypotheses, we find that, upon letting ε tend to 0 and R tend to ∞,

$$\begin{aligned} u(s) &= \frac{\Gamma(s+1)(e^{-\pi i s} - e^{\pi i s})}{2\pi i}\int_0^\infty f(x) x^{-s-1}\, dx \\ &= -\frac{\Gamma(s+1)\sin(\pi s)}{\pi}\int_0^\infty f(x) x^{-s-1}\, dx. \end{aligned}$$

From this formula and our hypotheses, it follows that $u(s)$ is analytic in a neighborhood of $s = 0$. Calculating $u'(0)$ from the formula above, we deduce (2.6) to complete the proof.

Hardy [**11**] has written a fascinating paper on generalizations of Frullani's theorem. Included in the paper is a dazzling assortment of examples that are instances of (2.2) and that were developed by Hardy ad hoc. For example,

$$\int_0^\infty \left(\cos x - \frac{\sin x}{x}\right)\frac{dx}{x} = -1$$

and

$$\int_0^\infty \frac{\cos x - e^{-x^2}}{x}\, dx = -\frac{\gamma}{2},$$

where γ denotes Euler's constant. These two examples do not fall under the province of Theorem 2; Hafner's lemma is needed to justify the use of (2.2). Although Hardy discovered several generalizations of Frullani's theorem in [**11**] and other papers, it is interesting that he, or anyone else, never discerned Ramanujan's beautiful generalization.

In the last portion of the second quarterly report, Ramanujan derives eight corollaries of his Master Theorem and employs two of them and analogues to establish the inversion theorem for Fourier cosine and sine transforms. We give those two corollaries needed for the cosine formulae.

COROLLARY 1.

$$\int_0^\infty \sum_{k=0}^\infty \frac{\phi(k)(-x)^k}{k!}\cos(nx)\, dx = \sum_{k=0}^\infty \phi(-2k-1)(-n^2)^k.$$

Ramanujan's Proof. Expand $\cos(nx)$ into its Maclaurin series, invert the order of integration and summation for this series, and then apply the Master Theorem to each term.

For a rigorous proof of Corollary 1, see Hardy's book [**14**, pp. 200–201]. For Ramanujan's proof of Corollary 2 as well as a rigorous proof, see [**6**].

COROLLARY 2.

$$\int_0^\infty \sum_{k=0}^\infty \phi(2k)(-x^2)^k\cos(nx)\, dx = \frac{\pi}{2}\sum_{k=0}^\infty \frac{\phi(-k-1)(-n)^k}{k!}.$$

The inversion theorem below is valid if f is of bounded variation on every finite interval in $[0, \infty)$ and f is absolutely integrable over $(0, \infty)$ [**24**, pp. 434–435].

THEOREM 3. *Let* $n > 0$. *If*

$$\int_0^\infty f(x)\cos(nx)\,dx = g(n), \tag{2.8}$$

then

$$\int_0^\infty g(x)\cos(nx)\,dx = \frac{\pi}{2}f(n). \tag{2.9}$$

Ramanujan's Proof. If

$$f(x) = \sum_{k=0}^{\infty} \frac{\phi(k)(-x)^k}{k!},$$

then by Corollary 1, we observe that (2.8) holds with

$$g(x) = \sum_{k=0}^{\infty} \phi(-2k-1)(-x^2)^k.$$

An application of Corollary 2 then yields (2.9).

Ramanujan concludes the second quarterly report with several well-known examples to illustrate Theorem 3.

3. The Third Quarterly Report. In the last of his quarterly reports, Ramanujan considers additional ramifications of his Master Theorem, but examines a few new topics as well. He first discusses the ordinary composition of functions and extends this concept to "the fractional order of functions."

Define $F^0(x) = x$, $F^1(x) = F(x)$, and $F^n(x) = F(F^{n-1}(x))$, for $n \geqslant 2$ and any function F. Of course, the range of F^{n-1} must be contained in the domain of F. In order to define "the fractional order" of a particular function F, Ramanujan needs to first find a general formula for $F^n(x)$ when n is a nonnegative integer. He then extends this formula to all real values of n.

To illustrate Ramanujan's ideas, let $F(x) = x^2 - 2$, $x \geqslant 2$. Set $x = y + (1/y)$. Then

$$F(x) = y^2 + y^{-2}, \quad F^2(x) = y^4 + y^{-4},$$

and, in general,

$$F^n(x) = y^{2^n} + y^{-2^n} = \left(\frac{x + \sqrt{x^2-4}}{2}\right)^{2^n} + \left(\frac{x - \sqrt{x^2-4}}{2}\right)^{2^n}.$$

Hence, for example,

$$F^{\operatorname{Log}3/\operatorname{Log}2}(x) = x^3 - 3x.$$

As a natural outgrowth of his study of the fractional composition of functions, Ramanujan briefly studies fractional differentiation. For each nonnegative integer n, let $D^n f(x) = f^{(n)}(x)$. Ramanujan assumes that there exists a unique, "natural" function of n passing through the points $D^0 f(x)$, $Df(x)$, $D^2 f(x), \ldots,$ and in this rather imprecise fashion defines fractional derivatives of f. Thus, for every real number n, Ramanujan defines $D^n e^{ax} = a^n e^{ax}$. Assuming that the kth fractional derivative $f^{(k)}(t)$ is well-defined, Ramanujan shows that it must satisfy the following relation.

THEOREM 4. *If* $n > 0$, *then*

$$\int_0^\infty x^{n-1} f^{(r)}(a-x)\,dx = \Gamma(n) f^{(r-n)}(a). \tag{3.1}$$

Ramanujan points out that Theorem 4 can actually be used to *define* the fractional derivative $f^{(k)}$. Let r be any nonnegative integer greater than k and let $n = r - k$. Thus, $n > 0$, and by (3.1),

$$\int_0^\infty x^{n-1} f^{(r)}(a-x)\,dx = \Gamma(n) f^{(k)}(a). \tag{3.2}$$

Since the left side of (3.2) has a definite meaning from elementary calculus, (3.2) can be used to define the fractional derivative $f^{(k)}(a)$ for any real number k. It is remarkable that (3.2) is precisely the same definition that Liouville [17], one of the founders of fractional calculus, gave for the fractional derivative in 1832.

We now present Ramanujan's amazingly simple, but formal proof.

Ramanujan's Proof. By Taylor's theorem and the Master Theorem,

$$\begin{aligned}\int_0^\infty x^{n-1} f^{(r)}(a-x)\,dx &= \int_0^\infty x^{n-1} \sum_{k=0}^\infty \frac{f^{(r+k)}(a)(-x)^k}{k!}\,dx \\ &= \Gamma(n) f^{(r-n)}(a).\end{aligned}$$

The last portion of the third report is devoted primarily to various types of integral formulas and transforms. Besides using his Master Theorem, Ramanujan also employs operational methods. We offer one such example which Ramanujan prefaces by declaring that "If a result is true *only* for real values of a quantity (say a), then the result got by using the operator for a is true only if when the new function can be expressed in terms of the original function..." As an illustration, Ramanujan considers

$$\int_0^\infty \frac{\cos(ax)}{x^2+1}\,dx = \frac{\pi}{2} e^{-a},$$

which is valid only for (nonnegative) real values of a. Multiplying both sides by $\phi(0)$ and replacing a by the differential operator D, we get

$$\int_0^\infty \frac{\cos(Dx)}{x^2+1}\phi(0)\,dx = \frac{\pi}{2} e^{-D}\phi(0). \tag{3.3}$$

Now

$$\begin{aligned}\cos(Dx)\phi(0) &= \frac{e^{iDx}+e^{-iDx}}{2}\phi(0) \\ &= \frac{1}{2}\sum_{k=0}^\infty \frac{(iDx)^k + (-iDx)^k}{k!}\phi(0) \\ &= \frac{1}{2}\sum_{k=0}^\infty \frac{(ix)^k + (-ix)^k}{k!}\phi^{(k)}(0) = \frac{\phi(ix)+\phi(-ix)}{2}.\end{aligned}$$

Similarly, $e^{-D}\phi(0) = \phi(-1)$. Thus, (3.3) can be written in the form

$$\int_0^\infty \frac{\phi(ix)+\phi(-ix)}{2(x^2+1)}\,dx = \frac{\pi}{2}\phi(-1),$$

which Ramanujan claims is valid only when $\phi(ix) + \phi(-ix)$ can be written as a linear combination of cosines.

For many additional examples see Hardy's book [**14**, Chapter 11] and our paper [**4**].

Conclusion. Although Ramanujan gave only formal proofs for most of the results found in the quarterly reports, it is remarkable that all of his theorems are correct (if hypotheses are appended) and that all of his examples are likewise valid. Despite the shortage of hypotheses for his theorems, Ramanujan possessed an uncanny ability in discerning when an application or formula

was true and when it was not. His creative capacity to produce wondrous formulas is not only unsurpassed in this century, when beautiful formulas are no longer at their zenith, but perhaps in the history of mathematics as well. The quarterly reports may contain a message for contemporary mathematicians. We might allow our thoughts to occasionally escape from the chains of rigor and, in their freedom, to discover new pathways through the forest.

References

1. G. E. Andrews, An introduction to Ramanujan's "lost" notebook, this MONTHLY, 86 (1979) 89–108.
2. R. Askey, Ramanujan's extensions of the gamma and beta functions, this MONTHLY, 87 (1980) 346–359.
3. B. C. Berndt, Ramanujan's notebooks, Math. Mag., 51 (1978) 147–164.
4. B. C. Berndt, Ramanujan's quarterly reports, Bull. London Math. Soc. (to appear).
5. B. C. Berndt, R. J. Evans, and B. M. Wilson, Chapter 3 of Ramanujan's second notebook, Advances in Math., to appear.
6. B. C. Berndt and B. M. Wilson, Chapter 4 of Ramanujan's second notebook, Proc. Royal Soc. Edinburgh, 89A (1981) 87–109.
7. G. S. Carr, Formulas and Theorems in Pure Mathematics, 2nd ed., Chelsea, New York, 1970.
8. F. Carlson, Sur une classe de séries de Taylor, Thesis, Upsala, 1914.
9. L. Euler, De serie Lambertina plurimisque eius insignibus proprietatibus, Opera Omnia, Serie 1, Bd. 6, B. G. Teubner, Leipzig, 1921, pp. 350–359.
10. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, 4th ed., Academic Press, New York, 1965.
11. G. H. Hardy, On the Frullanian integral $\int_0^\infty \frac{\{\phi(ax^m)-\psi(bx^n)\}(\log x)^p}{x}dx$, Quart. J. Math., 33 (1902) 113–144.
12. G. H. Hardy, Orders of Infinity, Cambridge University Press, London, 1910.
13. G. H. Hardy, The Indian mathematician Ramanujan, this MONTHLY, 44 (1937) 137–155.
14. G. H. Hardy, Ramanujan, Chelsea, New York, 1978.
15. J. L. Lagrange, Nouvelle méthode pour résoudre les équations littérales par le moyen des séries, Oeuvres, vol. 3, Gauthier-Villars, Paris, 1869, pp. 5–73.
16. J. H. Lambert, Observationes variae in mathesin puram, Opera Mathematica, vol. 1, Orell Füssli, Zürich, 1946, pp. 16–51.
17. J. Liouville, Mémoire sur quelques questions de géométrie et de mécanique, et sur un nouveau genre de calcul pour résoudre ces questions, J. l'Ecole Polytech., 13 (1832), XXIe cahier, pp. 1–69.
18. G. Pólya and G. Szegő, Problems and Theorems in Analysis, vol. 1, Springer-Verlag, Berlin, 1972.
19. K. G. Ramanathan, The unpublished manuscripts of Srinivasa Ramanujan, Current Sci., 50 (1981) 203–210.
20. S. Ramanujan, Some definite integrals, Mess. Math., 44 (1915) 10–18.
21. S. Ramanujan, Collected papers, Chelsea, New York, 1962.
22. S. Ramanujan, Notebooks (2 vols.), Tata Institute of Fundamental Research, Bombay, 1957.
23. R. A. Rankin, Ramanujan's manuscripts and notebooks, Bull. London Math. Soc., 14 (1982) 81–97.
24. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Clarendon Press, Oxford, 1939.
25. G. N. Watson, Ramanujan's notebooks, J. London Math. Soc., 6 (1931) 137–153.
26. B. Williamson, An elementary Treatise on the Integral Calculus, 7th ed., Longmans, Green, and Co., New York, 1896.

MISCELLANEA

112. The ideal mathematical preparation for a user of mathematics such as a biologist or a chemist should focus not on acquiring skills but on acquiring certain attitudes; the most important attitude is the courage to sit down and try to figure something out.

—Paraphrased from "A matter of opinion" by Mark Kac, in *Scientific Research*, 3 (1968) 30–32.

was true and when it was not. His creative capacity to produce wondrous formulas is not only unsurpassed in this century, when beautiful formulas are no longer at their zenith, but perhaps in the history of mathematics as well. The quarterly reports may contain a message for contemporary mathematicians. We might allow our thoughts to occasionally escape from the chains of rigor and, in their freedom, to discover new pathways through the forest.

References

1. G. E. Andrews, An introduction to Ramanujan's "lost" notebook, this MONTHLY, 86 (1979) 89–108.
2. R. Askey, Ramanujan's extensions of the gamma and beta functions, this MONTHLY, 87 (1980) 346–359.
3. B. C. Berndt, Ramanujan's notebooks, Math. Mag., 51 (1978) 147–164.
4. B. C. Berndt, Ramanujan's quarterly reports, Bull. London Math. Soc. (to appear).
5. B. C. Berndt, R. J. Evans, and B. M. Wilson, Chapter 3 of Ramanujan's second notebook, Advances in Math., to appear.
6. B. C. Berndt and B. M. Wilson, Chapter 4 of Ramanujan's second notebook, Proc. Royal Soc. Edinburgh, 89A (1981) 87–109.
7. G. S. Carr, Formulas and Theorems in Pure Mathematics, 2nd ed., Chelsea, New York, 1970.
8. F. Carlson, Sur une classe de séries de Taylor, Thesis, Upsala, 1914.
9. L. Euler, De serie Lambertina plurimisque eius insignibus proprietatibus, Opera Omnia, Serie 1, Bd. 6, B. G. Teubner, Leipzig, 1921, pp. 350–359.
10. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, 4th ed., Academic Press, New York, 1965.
11. G. H. Hardy, On the Frullanian integral $\int_0^\infty \frac{\{\phi(ax^m)-\psi(bx^n)\}(\log x)^p}{x}dx$, Quart. J. Math., 33 (1902) 113–144.
12. G. H. Hardy, Orders of Infinity, Cambridge University Press, London, 1910.
13. G. H. Hardy, The Indian mathematician Ramanujan, this MONTHLY, 44 (1937) 137–155.
14. G. H. Hardy, Ramanujan, Chelsea, New York, 1978.
15. J. L. Lagrange, Nouvelle méthode pour résoudre les équations littérales par le moyen des séries, Oeuvres, vol. 3, Gauthier-Villars, Paris, 1869, pp. 5–73.
16. J. H. Lambert, Observationes variae in mathesin puram, Opera Mathematica, vol. 1, Orell Füssli, Zürich, 1946, pp. 16–51.
17. J. Liouville, Mémoire sur quelques questions de géométrie et de mécanique, et sur un nouveau genre de calcul pour résoudre ces questions, J. l'Ecole Polytech., 13 (1832), XXIe cahier, pp. 1–69.
18. G. Pólya and G. Szegö, Problems and Theorems in Analysis, vol. 1, Springer-Verlag, Berlin, 1972.
19. K. G. Ramanathan, The unpublished manuscripts of Srinivasa Ramanujan, Current Sci., 50 (1981) 203–210.
20. S. Ramanujan, Some definite integrals, Mess. Math., 44 (1915) 10–18.
21. S. Ramanujan, Collected papers, Chelsea, New York, 1962.
22. S. Ramanujan, Notebooks (2 vols.), Tata Institute of Fundamental Research, Bombay, 1957.
23. R. A. Rankin, Ramanujan's manuscripts and notebooks, Bull. London Math. Soc., 14 (1982) 81–97.
24. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Clarendon Press, Oxford, 1939.
25. G. N. Watson, Ramanujan's notebooks, J. London Math. Soc., 6 (1931) 137–153.
26. B. Williamson, An elementary Treatise on the Integral Calculus, 7th ed., Longmans, Green, and Co., New York, 1896.

MISCELLANEA

112. The ideal mathematical preparation for a user of mathematics such as a biologist or a chemist should focus not on acquiring skills but on acquiring certain attitudes; the most important attitude is the courage to sit down and try to figure something out.

—Paraphrased from "A matter of opinion" by Mark Kac, in *Scientific Research*, 3 (1968) 30–32.

What do these men have in common? See p. 570.

SOLUTIONS OF ADVANCED PROBLEMS

Flexible Linkages of Polygons

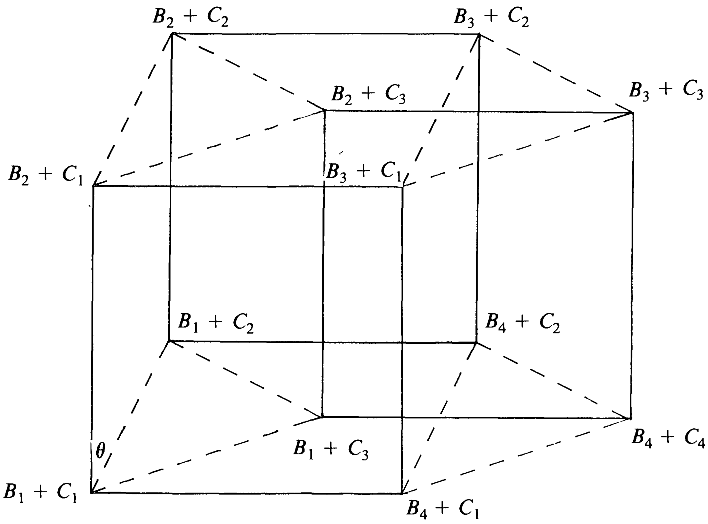
6367 [1981, 711]. *Proposed by A. Ehrenfeucht and J. Mycielski, University of Colorado.*

Let A be a finite collection of distinct but possibly overlapping regular n-gons of the same size on the plane such that every vertex of every n-gon of A is a vertex of exactly two n-gons of A.

(a) Construct a collection A of $2n$ n-gons such that, even if the n-gons are rigid, A is flexible.

(b)* For which n is a rigid A possible?

Solution to part (a) *by Kit Hanes, Eastern Washington University.* We establish a more general result. Let B be any n-gon with vertices $B_1,\ldots, B_n$ and let C be any m-gon with vertices $C_1,\ldots, C_m$. Let θ be the angle between $\overline{B_1B_2}$ and $\overline{C_1C_2}$. Then, for $i \in \{1,\ldots, n\}$ let $\hat{C}_i = B_i + C$, and for $j \in \{1,\ldots, m\}$ let $\hat{B}_j = B + C_j$. Let $A = \{\hat{C}_1,\ldots, \hat{C}_n, \hat{B}_1,\ldots, \hat{B}_m\}$. A is a collection of $n + m$ polygons each of which is a translate of either B or C. Any vertex of any polygon in A is a point $B_i + C_j$, which is a vertex just of $\hat{B}_j$ and of $\hat{C}_i$. This construction is valid independently of θ; as θ varies A flexes.



Also solved by Michael Goldberg and the proposers.

Part (b). The proposers state that they know of only two cases when A as defined in the problem is rigid, namely $n = 2$ (a triangle) and $n = 4$ (a configuration of 12 squares inside a regular 12-gon whose outer edges form the 12-gon).

ANSWER TO PHOTOS ON PAGE 517

They are all outstanding mathematicians, and they were all presidents of the A.M.S. Here are their names, followed by the years of their presidency. Top row: J. von Neumann (1951–1952), R. L. Wilder (1955–1956); bottom row: J. L. Doob (1963–1964), Oscar Zariski (1969–1970).

THE EULER ϕ-FUNCTION IN THE GAUSSIAN INTEGERS

JAMES T. CROSS
Department of Mathematics, The University of the South, Sewanee, TN 37375

1. Introduction. The set of all complex numbers $a + bi$, where a and b are integers, forms an integral domain with the usual complex number operations. This domain, the *Gaussian Integers*, which we denote by G, includes Z, the domain of *rational* (*ordinary*) *integers*, and behaves like Z in many respects.

Let β denote a nonzero Gaussian integer and $G/(\beta)$ the quotient ring of G (mod β). It is fun to discover that these rings furnish accessible and interesting examples of familiar concepts including equivalence classes, fields, cyclic and noncyclic groups, and the Euler ϕ-function. In this section we review the ϕ-function, extend it to G, and discuss some questions whose answers are well known in Z and which we intend to study in G.

Let n be a nonzero member of Z and let $Z/(n)$ denote the quotient ring of Z (mod n). If $n > 0$, then

$$Z/(n) = \{[0], [1], [2], \ldots, [n-1]\},$$

where the brackets denote *equivalence classes*. The rings $Z/(n)$ and $Z/(-n)$ are identical, so that we may confine our attention to positive n. The *units* (those members having multiplicative inverses) of this ring form a multiplicative group which is denoted by $\Phi_Z(n)$ and the Euler ϕ-function $\phi_Z(n)$ is defined to be the order of this group. A simple argument in elementary number theory proves that the member $[k]$ of $Z/(n)$ is a unit if and only if $(k, n) = 1$; that is, if and only if k and n are *relatively prime*. Thus for $n > 1$, $\phi_Z(n)$ is the number of positive integers less than n and prime to n. If $\Phi_Z(n)$ is cyclic and $[k]$ generates this group, then k is called a *primitive root* for n. Simple construction of multiplication tables leads to the conclusion that primitive roots exist for $n = 1, 2, 3, 4, 5, 6$, and 7. However no primitive root exists for 8 since $\Phi_Z(8) = \{[1], [3], [5], [7]\}$ has members of orders 1 and 2 only.

The determination of those integers that have primitive roots is a popular undertaking in number theory. There it is proved that n has primitive roots if and only if $n = 2$ or 4, or n is a power of an odd prime or twice a power of an odd prime. In fact, a great deal more is established. Let the symbol $\simeq$ denote *isomorphism* and Z_n denote the (cyclic) additive group of $Z/(n)$. Then for $m > 1$, the structure of $\Phi_Z(m)$ is given as follows [**1**, pages 46–51]:

$\Phi_Z(2) \simeq Z_1$,

$\Phi_Z(4) \simeq Z_2$,

$\Phi_Z(2^n) \simeq Z_2 \times Z_{2^{n-2}}$ if $n > 2$,

$\Phi_Z(p^n) \simeq Z_{p^n - p^{n-1}}$ if p is an odd prime,

Φ_Z is multiplicative in the sense that $\Phi_Z(mn) \simeq \Phi_Z(m) \times \Phi_Z(n)$ if $(m, n) = 1$.
(The structure of $\Phi_Z(m)$ for arbitrary m is gotten through its prime factorization.)

The Euler ϕ-function extends naturally to G. Since Z is contained in G and the units of $G/(\beta)$ also form a multiplicative group, we let $\Phi_G(\beta)$ denote this group and let $\phi_G(\beta)$ denote its order. The following questions arise immediately:

James T. Cross: I received the Ph.D. from the University of Tennessee where Professor Eckford Cohen directed my work. I majored in mathematics at Brown University, took an M.S. in applied mathematics at Harvard, and have been at Sewanee since 1955. The problem discussed in this paper occurred to me as a result of some "primitive root" studies done by two of my senior honors students. My extracurricular activities are extensive and intensive (too much so). I mention two: mini-farming and ping-pong.

1. If p is an odd rational prime, then is $\Phi_G(p^n)$ still cyclic in this larger setting?
2. More generally, which Gaussian integers have primitive roots?
3. Can we determine the structure of $\Phi_G(\beta^n)$, where β is prime in G?
4. If the answer to question 3 is yes, can we extend our results to an arbitrary Gaussian integer through its prime factorization?

Question 1 can be settled quickly in the negative by determining the orders of the members of $\Phi_G(3^2)$. As we shall show later in Example 4, a complete and nonrepeating set of representatives for the members (equivalence classes) of this group is the set of all $a + bi$, where a and b range independently from 0 through 8 and at least one of a and b is prime to 3. The 72 members of this group, and their orders, are tabulated in Fig. 1. This table is taken from an unpublished paper [**2**], written by my student, Michael S. Crowe, who answered questions 1 and 2.

	$[i]$	$[2i]$		$[4i]$	$[5i]$		$[7i]$	$[8i]$
	4	12		12	12		12	4
$[1]$	$[1+i]$	$[1+2i]$	$[1+3i]$	$[1+4i]$	$[1+5i]$	$[1+6i]$	$[1+7i]$	$[1+8i]$
1	24	24	3	24	24	3	24	24
$[2]$	$[2+i]$	$[2+2i]$	$[2+3i]$	$[2+4i]$	$[2+5i]$	$[2+6i]$	$[2+7i]$	$[2+8i]$
6	24	8	6	24	24	6	8	24
	$[3+i]$	$[3+2i]$		$[3+4i]$	$[3+5i]$		$[3+7i]$	$[3+8i]$
	12	12		12	12		12	12
$[4]$	$[4+i]$	$[4+2i]$	$[4+3i]$	$[4+4i]$	$[4+5i]$	$[4+6i]$	$[4+7i]$	$[4+8i]$
3	24	24	3	24	24	3	24	24
$[5]$	$[5+i]$	$[5+2i]$	$[5+3i]$	$[5+4i]$	$[5+5i]$	$[5+6i]$	$[5+7i]$	$[5+8i]$
6	24	24	6	24	24	6	24	24
	$[6+i]$	$[6+2i]$		$[6+4i]$	$[6+5i]$		$[6+7i]$	$[6+8i]$
	12	12		12	12		12	12
$[7]$	$[7+i]$	$[7+2i]$	$[7+3i]$	$[7+4i]$	$[7+5i]$	$[7+6i]$	$[7+7i]$	$[7+8i]$
3	24	8	3	24	24	3	8	24
$[8]$	$[8+i]$	$[8+2i]$	$[8+3i]$	$[8+4i]$	$[8+5i]$	$[8+6i]$	$[8+7i]$	$[8+8i]$
2	24	24	6	24	24	6	24	24

FIG. 1.
The elements of $\Phi_G(3^2)$ and their orders.

In the table, for which the computations were done by computer, square brackets denote equivalence classes while the numbers without brackets indicate orders. For example, the order of $[4 + 7i]$ is 24, the order of [5] is 6, and the order of $[1 + 3i]$ is 3. These orders are easy to determine though the calculations are tedious. For example, the order of $[1 + 3i]$ is 3 because

$$(1 + 3i)^3 = 1 + 9i + 27i^2 + 27i^3 \equiv 1 \pmod 9,$$

while

$$(1 + 3i)^2 = 1 + 6i + 9i^2 \not\equiv 1 \pmod 9.$$

Since the highest order observed in the table is 24, it is clear that there exist odd primes for which the answer to question 1 is no. (We will show in Example 7 below that $\Phi_G(3^2) \simeq Z_3 \times Z_3 \times Z_8$.)

Our objective in this paper is to answer questions 3 and 4, obtaining Crowe's answer to question 2 as a by-product. Our methods are direct and constructive:

(i). We identify the primes in G. Fortunately, we have to consider only three types.

(ii). For β a prime in G, we find equivalence class representatives of $G/(\beta^n)$ and its units group, $\Phi_G(\beta^n)$.

(iii). For a power of each type of prime, we determine the structure of the corresponding units group, answering question 3.

(iv). We show that Φ_G is multiplicative. The answer to question 4 will then be at hand.

2. The Primes in G. Our principal goal in this section is to identify the primes in G; in conjunction with this program we also give some relevant facts concerning the Gaussian Integers. The reader can find these topics discussed in [**1**, pp. 82–112] or [**3**, pp. 246–260]:

For $\beta = a + bi$ in G, the norm $N(\beta)$ of β is defined by $N(\beta) = (a + bi)(a - bi) = a^2 + b^2$. Note that $N(\beta)$ is in Z. The norm of a product equals the product of the norms. If β divides γ in G, then $N(\beta)$ divides $N(\gamma)$ in Z.

If μ is in G, then μ is a unit if and only if $N(\mu) = 1$. It follows that the units in G are $1, -1, i$, and $-i$.

If q is a positive prime in Z and $q \equiv 1 \pmod 4$, then $q = \pi\bar{\pi}$ for some π in G, where $\bar{\pi}$ is the complex conjugate of π. Here π and $\bar{\pi}$ are prime in G and they are not associates. (Members β and γ in G are *associates*, which we denote by $\beta \sim \gamma$, if $\beta = \mu\gamma$ for some unit μ in G.)

EXAMPLE 1. In G, $5 = (1 + 2i)(1 - 2i)$. If, for example, $1 + 2i = \beta\gamma$ in G, then $N(1 + 2i) = 5 = N(\beta)\, N(\gamma)$ in Z. Then $N(\beta) = 1$ or $N(\gamma) = 1$, so that one of the factors is a unit and the factorization is trivial. Thus $1 + 2i$ is prime; similarly, so is $1 - 2i$. To see that they are not associates, we need only try to get from one to the other by means of the four units.

If p is a positive prime in Z and $p \equiv 3 \pmod 4$, then p is prime in G.

EXAMPLE 2. Let $p = 3$. If $3 = \beta\gamma$ in G, then $N(3) = 9 = N(\beta)\, N(\gamma)$ in Z. Then either $N(\beta) = N(\gamma) = 3$ or the norm of one of the factors is 1. Since $a^2 + b^2 = 3$ is not solvable in Z, the first alternative must be discarded. Then one of β and γ is a unit, the factorization of 3 is trivial, and 3 is prime in G.

In G, $2 = (1 + i)(1 - i)$, where these factors are primes. However, they are associates since $1 - i = -i(1 + i)$. Then $2 = \mu(1 + i)^2$, where μ is a unit in G, and we see that 2 is a power of a prime in G. If α denotes $1 + i$, then $\alpha^2 \sim 2$, $\alpha^3 \sim 2\alpha$, $\alpha^4 \sim 4, \dots, \alpha^{2m} \sim 2^m$, and $\alpha^{2m+1} \sim 2^m\alpha$.

The primes described above (α and primes of the types π and p) are, together with their associates, the only primes in G.

Factorization into primes in G is unique in the following sense: two factorizations may look different, but we can get from one to the other by multiplying by units. For example, as we saw above, $2 = \alpha\bar{\alpha} = -i\alpha^2$. This phenomenon also occurs in Z: $6 = (3)(2) = (-3)(-2)$.

3. The Equivalence Classes in $G/(\beta^n)$ and $\Phi_G(\beta^n)$, Where β Is Prime in G. Now that we have identified the primes in G, we want to be able to raise them to powers and find representatives for the equivalence classes of the corresponding quotient rings and units groups. Theorems 1 and 2 below address this problem. Following the theorems we consider some examples.

In the remainder of the paper we will continue with the following notation: q and p will denote positive primes in Z satisfying $q \equiv 1 \pmod 4$ and $p \equiv 3 \pmod 4$, π will denote a *prime factor of q in G*, and α will denote $1 + i$.

THEOREM 1. *The equivalence classes of G* modulo *a power of a prime are given as follows*:

1. $G/(\pi^n) = \{[a]\colon 0 \leqslant a \leqslant q^n - 1\}$,
2. $G/(p^n) = \{[a + bi]\colon 0 \leqslant a \leqslant p^n - 1$ *and* $0 \leqslant b \leqslant p^n - 1\}$,
3. $G/(\alpha^{2m}) = \{[a + bi]\colon 0 \leqslant a \leqslant 2^m - 1$ *and* $0 \leqslant b \leqslant 2^m - 1\}$,
4. $G/(\alpha^{2m+1}) = \{[a + bi]\colon 0 \leqslant a \leqslant 2^{m+1} - 1$ *and* $0 \leqslant b \leqslant 2^m - 1\}$.

In the statements of this theorem, as well as in the examples to be considered below, we intend to imply that the given sets of representatives are complete and nonrepeating.

Proof. First we observe that $G/(\alpha^{2m}) = G/(2^m)$ and $G/(\alpha^{2m+1}) = G/(2^m\alpha)$ because $\alpha^{2m} \sim 2^m$. Now if $a + bi \equiv c + di \pmod{\alpha^{2m}}$, then 2^m divides both $a - c$ and $b - d$, so that the classes of (3) are distinct. A similar argument applies to the classes in (2). If $[a] = [b]$ in $G/(\pi^n)$, then π^n divides $a - b$. Let $\pi^n\gamma = a - b$ for some γ in G. Then taking complex conjugates, we get $\bar{\pi}^n = \overline{a - b} = a - b$, so that $\bar{\pi}^n$ also divides $a - b$. Since π and $\bar{\pi}$ are not associates, $\pi^n\bar{\pi}^n = q^n$

divides $a - b$ implying that the classes in (1) are distinct. If $[a + bi] = [c + di]$ in $G/(\alpha^{2m+1}) = G/(2^m\alpha)$, then $2^m\alpha$ divides $a - c + (b - d)i$. Then 2^m divides $b - d$; since each is less than 2^m, $b = d$. Then $2^m\alpha$ divides $a - c$. Let $a - c = 2^m k$, where k is in Z because the only rational members of G are integers. Then α divides k so that $N(\alpha) = 2$ divides $N(k) = k^2$. Then 2 divides k, so that 2^{m+1} divides $a - c$ and the classes in (4) are distinct.

Now let $\beta = x + yi$ be in G. Reducing x and y by multiples of 2^m gets β in one of the classes in (3). Reducing by multiples of p^n, we get β in one of the classes of (2). Reducing by multiples of 2^{m+1} yields $\beta \equiv c + di \pmod{2^{m+1}}$, where each of c and d is nonnegative and less than 2^{m+1}. If $d < 2^m$, we have β in one of the classes of (4). If $d \geqslant 2^m$, then we subtract and add $2^m\alpha$, getting

$$c + di = (c - 2^m) + (d - 2^m)i + 2^m\alpha.$$

Then

$$\beta \equiv (c - 2^m) + (d - 2^m)i \pmod{\alpha^{2m+1}},$$

where $0 \leqslant d - 2^m < 2^m$. Then $c - 2^m$ can be reduced by multiples of 2^{m+1}, getting β in one of the classes in (4). Now we show that i belongs to one of the classes in (1). Let $\pi^n = a - bi$, so that $bi \equiv a \pmod{\pi^n}$. Now q is prime to b, for if q divides b, then both π and $\bar{\pi}$ divide b. Then π divides a. But this implies that q divides a. Then q divides π^n; this is absurd because $q = \pi\bar{\pi}$. Thus $(q, b) = 1$ and the congruence, $zb \equiv 1 \pmod{q^n}$ is solvable in Z. Then the congruence, $bi \equiv a \pmod{q^n}$, yields $i \equiv za \pmod{\pi^n}$. Then za can be reduced by multiples of p^n to find i in one of the given classes. Since each of x, y, and i belongs to one of the classes of (1), then so does $\beta = x + yi$. □

This theorem implies that $G/(\pi^n)$ has q^n members, $G/(p^n)$ has p^{2n} members, and $G/(\alpha^n)$ has 2^n members. These facts are special cases of the well-known result [**4**, page 54] that the order of $G/(\beta)$ is $N(\beta)$.

Now we are ready to identify the units of the rings whose members are specified in Theorem 1.

THEOREM 2. *Let $[a]$ be in $G/(\pi^n)$. Then $[a]$ is a unit if and only if $(q, a) = 1$. Let $[a + bi]$ be in $G/(p^n)$. Then $[a + bi]$ is a unit if and only if at least one of a and b is prime to p. Let $[a + bi]$ be in $G/(\alpha^n)$. Then $[a + bi]$ is a unit if and only if $a \not\equiv b$* (mod 2).

Proof. Let β and γ be in G. Then $[\beta]$ is a unit in $G/(\gamma)$ if and only if $[\beta][\delta] = [1]$ in $G/(\gamma)$, for some δ in G. Then $[\beta]$ is a unit if and only if $\beta\delta \equiv 1 \pmod{\gamma}$; that is, if and only if $\beta\delta + \eta\gamma = 1$ for some η in G. Thus $[\beta]$ is a unit in $G/(\gamma)$ if and only if β is prime to γ. It follows that $[a]$ in $G/(\pi^n)$ is a unit if and only if π does not divide a, but π does not divide a if and only if q does not divide a. Next, $[a + bi]$ in $G/(p^n)$ is a unit if and only if p does not divide $a + bi$; p does not divide $a + bi$ if and only if p is prime to at least one of a and b. Finally, $[a + bi]$ in $G/(\alpha^n)$ is a unit if and only if α does not divide $a + bi$. But

$$(a + bi)/\alpha = (a + bi)/(1 + i) = (a + b)/2 + i(b - a)/2.$$

This quotient is a Gaussian integer if and only if $a \equiv b \pmod 2$. Thus α does not divide $a + bi$ if and only if $a \not\equiv b \pmod 2$. □

EXAMPLE 3. Let $\pi = 1 + 2i$. We know that π is prime and $\pi\bar{\pi} = 5$. Theorems 1 and 2 assert that

$$G/(\pi^2) = \{[0], [1], [2], \ldots, [24]\}$$

and $[a]$ in $G/(\pi^2)$ is a unit if and only if 5 does not divide a. It may be instructive to find the class to which i belongs. Now $\pi^2 = -3 + 4i$, so that $4i \equiv 3 \pmod{\pi^2}$. Multiplying this congruence through by 19, we get $76i \equiv 57 \pmod{\pi^2}$. But since π^2 divides 25, this congruence reduces to $i \equiv 7 \pmod{\pi^2}$. We can check this result directly: $(7 - i)/(-3 + 4i) = -1 - i$, so that π^2 divides $7 - i$ in G. Thus $i \equiv 7 \pmod{\pi^2}$, and i belongs to [7].

In this example the reader should convince himself that $\Phi_G((1 + 2i)^2)$ is isomorphic to $\Phi_Z(25)$.

Since $\Phi_Z(25)$ is cyclic, $(1 + 2i)^2 = -3 + 4i$ has primitive roots. In fact, 2 serves as such a root.

EXAMPLE 4. By Theorems 1 and 2, $\Phi_G(3^2) = \Phi_G(9) =$

$$\{[a + bi] : 0 \leqslant a \leqslant 8, 0 \leqslant b \leqslant 8, \text{and 3 is prime to at least one of } a \text{ and } b\}.$$

Note that $\Phi_Z(9)$ is included isomorphically in $\Phi_G(9)$. Thus $\Phi_Z(9) = \{[1], [2], [4], [5], [7], [8]\}$ can be identified in an obvious way with $\{[1], [2], [4], [5], [7], [8]\}$ in $\Phi_G(9)$.

EXAMPLE 5. By Theorems 1 and 2,

$$\begin{aligned}\Phi_G(\alpha^5) &= \Phi_G(4\alpha)\\ &= \{[1], [3], [5], [7], [i], [2 + i], [4 + i], [6 + i], [1 + 2i], [3 + 2i],\\ &\qquad [5 + 2i], [7 + 2i], [3i], [2 + 3i], [4 + 3i], [6 + 3i]\}.\end{aligned}$$

Note that $\Phi_Z(8) = \{[1], [3], [5], [7]\}$ is included isomorphically in $\Phi_G(\alpha^5)$. Since $\Phi_Z(8)$ is not cyclic and since any subgroup of a cyclic group is cyclic, α^5 has no primitive root.

Let us be more ambitious and find the structure of $\Phi_G(\alpha^5)$. Let H denote the subgroup generated by $[1 + 2i]$, and let K and I denote the subgroups generated by $[5]$ and $[i]$, respectively. Then

$$H = \{[1], [1 + 2i]\}, \quad K = \{[1], [5]\},$$

and

$$I = \{[1], [i], [-1], [-i]\} = \{[1], [i], [7], [4 + 3i]\}.$$

Now $H \times K = \{[1], [5], [1 + 2i], [5 + 2i]\}$. Since $H \times K$ intersects I only at the identity and since each of $H \times K$ and I has order 4, the order of $H \times K \times I$ is $16 = \phi_G(\alpha^5)$. Then

$$\Phi_G(\alpha^5) = H \times K \times I \simeq Z_2 \times Z_2 \times Z_4.$$

Now, using Theorems 1 and 2, we can count the members of the units groups and find the following results:

$$\begin{aligned}\phi_G(\pi^n) &= \phi_Z(q^n) = q^n - q^{n-1} = q^{n-1}(q - 1),\\ \phi_G(p^n) &= p^{2n} - p^{2n-2} = p^{2n-2}(p^2 - 1),\\ \phi_G(\alpha^n) &= 2^n - 2^{n-1} = 2^{n-1}.\end{aligned}$$

The reader is invited to verify these results in Examples 3–5.

In our program outlined in (i)–(iv) of Section 1, we have now reached (iii). We are ready to find the structures of the units groups that we have been discussing.

4. The Structure of $\Phi_G(\pi^n)$. We show that Example 3 is typical; that is, $\Phi_G(\pi^n)$ is cyclic.

THEOREM 3. $\Phi_G(\pi^n) \simeq Z_{q^n - q^{n-1}}$.

Proof. By Theorem 2,

$$\Phi_G(\pi^n) = \{[a]; 1 \leqslant a < q^n \text{ and } (q, a) = 1\}.$$

Formally, $\Phi_G(\pi^n)$ looks like $\Phi_Z(q^n)$, and, in fact, the map

$$[a] \text{ in } \Phi_Z(q^n) \to [a] \text{ in } \Phi_G(\pi^n)$$

is an isomorphism onto $\Phi_G(\pi^n)$. Thus, $\Phi_G(\pi^n)$ behaves algebraically like $\Phi_Z(q^n)$. Since $\Phi_Z(q^n) \simeq Z_{q^n - q^{n-1}}$, the proof is complete. While this theorem may be pleasing, it is not very exciting. Fortunately we encounter more interesting ones below.

5. Some Preliminary Ideas Concerning $\Phi_G(\alpha^n)$ and $\Phi_G(p^n)$. We dealt in Section 4 with one type of prime in G. The units groups corresponding to the other two types are generally not cyclic

as we saw in the Introduction and in Example 5. We showed in Example 5 that $\Phi_G(\alpha^5) = H \times K \times I$. Let us see whether we may expect a similar structure when α is raised to an even power.

EXAMPLE 6. By Theorems 1 and 2,

$$\Phi_G(\alpha^6) = \Phi_G(8) = \{[a + bi]: 0 \leqslant a \leqslant 7, 0 \leqslant b \leqslant 7, \text{and } a \not\equiv b \pmod 2\}.$$

Again let H, K, and I be generated by $[1 + 2i]$, $[5]$, and $[i]$, respectively. Then

$$H = \{[1], [1 + 2i], [5 + 4i], [5 + 6i]\},$$
$$K = \{[1], [5]\}, \text{and } I = \{[1], [i], [7], [7i]\}.$$

Then $K \times I = \{[1], [i], [7], [7i], [5], [5i], [3], [3i]\}$, so that $(K \times I) \cap H = \{[1]\}$. Then the order of $H \times K \times I$ is $32 = \phi_G(\alpha^6)$. Then again $\Phi_G(\alpha^6) = H \times K \times I$.

Examples 5 and 6 lead us to conjecture that $\Phi_G(\alpha^n) = H \times K \times I$, where these subgroups are generated by $[1 + 2i]$, $[5]$, and $[i]$, respectively. We delay checking out this conjecture in order to motivate another one concerning $\Phi_G(p^n)$. This conjecture involves a subgroup generated by $[1 + pi]$, and we shall study this subgroup and the subgroup H simultaneously.

Now $\phi_G(3^2) = 72$, $\phi_G(3^3) = 648$, and $\phi_G(7^2) = 2352$; it is clear that ideas concerning the structure of $\Phi_G(p^n)$ cannot be gotten by displaying multiplication tables. However, we can write a computer program to find the orders of the members of some of these groups and hope to observe something significant. The results of this effort are summarized below, where we have tabulated the highest orders observed and the order of $[1 + pi]$:

Group	Order	Highest order of an element	Order of $[1 + pi]$
$\Phi_G(3^2)$	$72 = 3^2(3^2 - 1)$	$24 = 3(3^2 - 1)$	3
$\Phi_G(3^3)$	$648 = 3^4(3^2 - 1)$	$72 = 3^2(3^2 - 1)$	9
$\Phi_G(7^2)$	$2352 = 7^2(7^2 - 1)$	$336 = 7(7^2 - 1)$	7
$\Phi_G(7^3)$	$115248 = 7^4(7^2 - 1)$	$2352 = 7^2(7^2 - 1)$	49
$\Phi_G(11^2)$	$14520 = 11^2(11^2 - 1)$	$1320 = 11(11^2 - 1)$	11

Note that in each case the highest order observed is $p^{n-1}(p^2 - 1)$ and the product of this number with p^{n-1} gives $\phi_G(p^n)$. This is a short table; the evidence is skimpy, but we are led to conjecture that $\Phi_G(p^n) = L \times K$, where L has order $p^{n-1}(p^2 - 1)$ and K has order p^{n-1}. This would imply that $\Phi_G(p^n) = H \times K \times R$, where each of H and K has order p^{n-1} and R has order $p^2 - 1$. We note also that $[1 + pi]$ appears to have order p^{n-1}.

EXAMPLE 7. Referring to Fig. 1, we let H, K, and R be the subgroups of $\Phi_G(9)$ generated by $[1 + 3i]$, $[4]$, and $[7 + 2i]$, respectively. Then

$$H = \{[1], [1 + 3i], [1 + 6i]\}, \quad K = \{[1], [4], [7]\},$$

and

$$R = \{[1], [7 + 2i], [i], [7 + 7i], [8], [2 + 7i], [8i], [2 + 2i]\}.$$

Since $H \cap K = \{[1]\}$, $H \times K$ has order 9, and since 8 is prime to 9, $H \times K$ intersects R only at the identity. Then $H \times K \times R$ has order $72 = \phi_G(3^2)$. Then $\Phi_G(3^2) = H \times K \times R$.

Now we recapitulate. Examples 5–7 and our display of orders have led to two conjectures:

(i). $\Phi_G(\alpha^n) = H \times K \times I$, where H, K, and I are generated by $[1 + 2i]$, $[5]$, and $[i]$, respectively.

(ii). $\Phi_G(p^n) = H \times K \times R$, where H is generated by $[1 + pi]$, K has order p^{n-1}, and R has order $p^2 - 1$.

We prove these conjectures in the next two sections (one of them requires some qualification).

At this point we study the subgroups denoted by H in (i) and (ii). Our principal tool is the following lemma.

LEMMA 1. *Let k denote a positive integer. Then,*

(A) $(1 + pi)^{p^k} = 1 + p^{k+1}i + p^{k+2}\gamma$, *where γ is in G.*
(B) $(1 + 2i)^{2^k} = 1 + 2^{k+1}(a + i) + 2^{k+2}\gamma$, *where γ is in G and a is an odd integer.*

Proof. Let β be in G, let r denote a prime in Z, let k be a positive integer, and let σ denote $(1 + \beta r)^{r^k}$. The reader can readily convince himself that it is probably true that

$$\sigma = 1 + \beta r^{k+1} + (\beta^2/2)(r^k - 1)r^{k+2} + \gamma r^{k+2},$$

where γ is some member of G. A proof based on the Binomial Theorem is somewhat tedious but not difficult. We omit the details. Now,

$$\text{if } r \neq 2, \text{ then } \sigma \equiv 1 + \beta r^{k+1} \pmod{r^{k+2}}.$$

If $r = 2$, then

$$\sigma \equiv 1 + 2^{k+1}\beta + \beta^2(2^k - 1)2^{k+1} \pmod{2^{k+2}}$$
$$\equiv 1 + 2^{k+1}\big((2^k - 1)\beta^2 + \beta\big) \pmod{2^{k+2}}.$$

The statements of the lemma follow from putting $\beta = i$ in these results. □

The following lemma gives the orders of the subgroups denoted by H in conjectures (i) and (ii).

LEMMA 2. *Let each of m and n be a positive integer greater than 1. Then*

A. *Let ρ denote $1 + pi$. The order of $[\rho]$ in $\Phi_G(p^n)$ is p^{n-1}.*
B. *Let δ denote $1 + 2i$. The order of $[\delta]$ in $\Phi_G(\alpha^{2m})$ and in $\Phi_G(\alpha^{2m+1})$ is 2^{m-1}.*

Proof. Put $k = n - 1$ in (A) and $k = m - 1$ in (B) of Lemma 1. Then

$$\rho^{p^{n-1}} \equiv 1 \pmod{p^n}$$

and

$$\delta^{2^{m-1}} \equiv 1 + 2^m(a + i) \pmod{2^{m+1}}.$$

But α divides $a + i$ since a is odd, and $\alpha^{2m} \sim 2^m$. Then

$$\delta^{2^{m-1}} \equiv 1 \pmod{\alpha^{2m+1}}.$$

Then the order of $[\rho]$ in $\Phi_G(p^n)$ is a divisor of p^{n-1}, and the order of $[\delta]$ in $\Phi_G(\alpha^{2m})$ and in $\Phi_G(\alpha^{2m+1})$ is a divisor of 2^{m-1}. It suffices to show now that $\rho^{p^{n-2}} \not\equiv 1 \pmod{p^n}$ and $\delta^{2^{m-2}} \not\equiv 1 \pmod{2^m}$. Put $k = n - 2$ in (A) and $k = m - 2$ in (B) of Lemma 1. Then

$$\rho^{p^{n-2}} \equiv 1 + p^{n-1}i \not\equiv 1 \pmod{p^n},$$

and

$$\delta^{2^{m-2}} \equiv 1 + 2^{m-1}(a + i) \quad \pmod{2^m}.$$

Since 2 does not divide $a + i$, $\delta^{2^{m-2}} \not\equiv 1 \pmod{2^m}$. □

As we implied in the conjectures, we intend to use the subgroups generated by $[1 + pi]$ and $[1 + 2i]$ and denoted by H as factors in direct products. For this purpose it will be necessary to have the following lemma which asserts that no member, except the identity, of these subgroups contains a real number or a pure imaginary number.

LEMMA 3. *Let each of m and n be a positive integer greater than 1. Let ρ denote $1 + pi$ and δ denote $1 + 2i$. No member, except $[1]$, of the subgroup of $\Phi_G(p^n)$ generated by $[\rho]$ and no member, except $[1]$, of the subgroup of $\Phi_G(\alpha^{2m})$ or of $\Phi_G(\alpha^{2m+1})$ generated by $[\delta]$ can be represented by a real number or a pure imaginary number.*

Proof. Let a complex number c be called *special* if c is in Z or $c = ic_1$ for some c_1 in Z. Suppose now that δ^b is congruent (mod 2^m) to a special number, for some b satisfying $0 < b < 2^{m-1}$. Let B denote the set of all such "bad" b. First, we show that 2^{m-2} is not in B. Put $k = m - 2$ in part (B) of Lemma 1. Then

$$\delta^{2^{m-2}} \equiv 1 + 2^{m-1}(a + i) \pmod{2^m}.$$

Now supposing $\delta^{2^{m-2}}$ congruent (mod 2^m) to a special number s, we have

$$1 + 2^{m-1}a - s + 2^{m-1}i \equiv 0 \pmod{2^m}.$$

If s is real, 2^m divides 2^{m-1}. If s is imaginary, 2^{m-1} divides 1. These contradictions guarantee that 2^{m-2} is not in B. Our technique for completing the proof is Fermat's "method of descent." Let L be the least member of B and let

$$2^{m-1} = Ld + r,$$

where $0 \leqslant r < L$. If $r = 0$, then $L = 2^t$ for some t satisfying $0 < t < m - 2$. (We showed $t \neq m - 2$.) Then

$$\delta^{2^{m-2}} = \delta^{L(2^{m-2-t})}.$$

Since L is in B, δ^L is congruent (mod 2^m) to a special number. It follows that $\delta^{L(2^{m-2-t})} = \delta^{2^{m-2}}$ is also congruent (mod 2^m) to such a number, which we proved impossible. Then $r \neq 0$. Since the order of $[\delta]$ is 2^{m-1},

$$[1] = [\delta^{Ld}][\delta^r] = [s][\delta^r] \text{ in } \Phi_G(2^m), \tag{1}$$

for some special number s. Let $s = x$ or $s = ix$ for some x in Z. Since $[s]$ is in the units group, x is odd. Then let y in Z satisfy $[yx] = [1]$ in $\Phi_Z(2^m)$. Then 2^m divides $yx - 1$ in Z, implying that 2^m divides $yx - 1$ in G, so that $[yx] = [1]$ in $\Phi_G(2^m)$. From (1) we get

$$[y] = [ys][\delta^r] \text{ in } \Phi_G(2^m). \tag{2}$$

If $s = x$, then $[\delta^r] = [y]$. If $s = ix$, we multiply (2) by $[-i]$, getting $[-iy] = [\delta^r]$. In either case, δ^r is congruent (mod 2^m) to a special number, contradicting the minimality of L. The set B is empty. This completes the proof of the lemma as it applies to α. Replacing 2 by p and m by n, we can use the same argument to prove the lemma as it applies to p. □

6. The Structure of $\Phi_G(p^n)$. We conjectured in Section 5 that $\Phi_G(p^n) = H \times K \times R$, where H is generated by $[1 + pi]$, K has order p^{n-1}, and R has order $p^2 - 1$. We established the useful properties of H. Now we go to work on K.

We saw in Example 4 that $\Phi_Z(9)$ is included in $\Phi_G(9)$ by an obvious isomorphism. More generally, the map,

$$[a] \text{ in } \Phi_Z(p^n) \to [a] \text{ in } \Phi_G(p^n),$$

is an isomorphism. We know that $\Phi_Z(p^n)$ is cyclic and that $\phi_Z(p^n) = p^{n-1}(p - 1)$. It follows that some [a] in $\Phi_Z(p^n)$ has order p^{n-1}. The isomorphism then implies that [a] has order p^{n-1} in $\Phi_G(p^n)$. We let K be the subgroup generated by [a]. Each member of K can be represented by a real number, so that $H \cap K = \{[1]\}$ and $H \times K$ has order p^{2n-2}.

Next, we turn to R. Since p is prime in G, $G/(p)$ is a field and $\Phi_G(p)$ is cyclic, being the multiplicative group of a finite field. The order of this group is $p^2 - 1$. Let $[\beta]$ generate $\Phi_G(p)$. Then $\beta^{p^2-1} \equiv 1 \pmod{p}$ so that $\beta^{p^2-1} = 1 + \gamma p$ for some γ in G. Then from the proof of Lemma 1,

$$\left(\beta^{p^2-1}\right)^{p^{n-1}} = 1 + \eta p^n$$

for some η in G. Hence,

$$\left(\beta^{p^{n-1}}\right)^{p^2-1} \equiv 1 \pmod{p^n},$$

so $[\beta^{p^{n-1}}]$ has order t, where t divides $p^2 - 1$. It follows that $\beta^{tp^{n-1}} \equiv 1 \pmod p$, whence $p^2 - 1$ divides tp^{n-1}. Then $p^2 - 1$ divides t, so that $t = p^2 - 1$ and $[\beta^{p^{n-1}}]$ has order $p^2 - 1$ in $\Phi_G(p^n)$. We let R denote the subgroup of $\Phi_G(p^n)$ generated by $[\beta^{p^{n-1}}]$.

Now since every member of $H \times K$ has order a power of p, $(H \times K) \cap R = \{[1]\}$, and the order of $H \times K \times R$ is $p^{2n-2}(p^2 - 1) = \phi_G(p^n)$. Our conjecture has checked out; $\Phi_G(p^n) = H \times K \times R$ and we have proved the following theorem:

THEOREM 4. $\Phi_G(p^n) \simeq Z_{p^{n-1}} \times Z_{p^{n-1}} \times Z_{p^2-1}$.

Our proof of this theorem is valid only for $n > 1$, but the theorem holds also for $n = 1$. In this case $Z_{p^{n-1}}$ is trivial.

REMARK. In Example 7 the subgroups K and R can be gotten without resort to Fig. 1: Just examine the 6 members of $\Phi_Z(9)$, finding that [4] has order 3, and let K be generated by [4] in $\Phi_G(9)$. Then write out the 8 members of $\Phi_G(3)$, getting $[1 + i]$ to be a generator. Then let R be generated by $[(1 + i)^3]$ in $\Phi_G(9)$.

7. The Structure of $\Phi_G(\alpha^n)$. We conjectured in Section 5 that $\Phi_G(\alpha^n) = H \times K \times I$, where H is generated by $[1 + 2i]$ while K and I are generated by [5] and $[i]$, respectively. However, in Lemmas 2 and 3 we assumed $m > 1$ so that for $n < 4$ we determine directly the structure of $\Phi_G(\alpha^n)$. Moreover, in $\Phi_G(\alpha^4) = \Phi_G(4)$, the subgroup generated by [5] is trivial. Therefore, we give the structure of $\Phi_G(\alpha^n)$ in two theorems, the first applicable for $n = 1, 2, 3$, or 4, while the second applies if $n \geqslant 5$.

THEOREM 5. *$\Phi_G(\alpha) \simeq Z_1$, $\Phi_G(\alpha^2) \simeq Z_2$, $\Phi_G(\alpha^3) \simeq Z_4$, and $\Phi_G(\alpha^4) \simeq Z_2 \times Z_4$.*

Proof. By Theorem 2,

$$\begin{aligned}
\Phi_G(\alpha) &= \{[1]\},\\
\Phi_G(\alpha^2) &= \{[1],[i]\},\\
\Phi_G(\alpha^3) &= \{[1],[3],[i],[2+i]\},\\
\Phi_G(\alpha^4) &= \{[1],[3],[i],[3i],[1+2i],[2+i],[2+3i],[3+2i]\}.
\end{aligned}$$

One can verify that $[i]$ generates $\Phi_G(\alpha^3)$ and that $\Phi_G(\alpha^4)$ is the direct product of the subgroups, $\{[1], [1 + 2i]\}$ and $\{[1], [i], [3], [3i]\}$. □

Now we assume $n \geqslant 5$ and prove conjecture (i) of Section 5. We investigated the subgroup H. The properties of K that are important for us are given in the following lemma.

LEMMA 4. *Let $n \geqslant 5$. The order of K is 2^{m-2} or 2^{m-1}, according as $n = 2m$ or $n = 2m + 1$. The member $[-1]$ of $\Phi_G(\alpha^n)$ is not in K.*

Proof. The maps

$$[a] \text{ in } \Phi_Z(2^m) \to [a] \text{ in } \Phi_G(\alpha^{2m})$$

and

$$[a] \text{ in } \Phi_Z(2^{m+1}) \to [a] \text{ in } \Phi_G(\alpha^{2m+1})$$

are isomorphisms. We noted one of them in Example 5. Now $\Phi_Z(2^m)$ and $\Phi_Z(2^{m+1})$ are direct products of the subgroups generated by $[-1]$ and [5]. (See [**1**, page 46].) The order of the subgroup generated by $[-1]$ is 2. Then the order of the subgroup generated by [5] is $\phi_Z(2^m)/2$ or $\phi_Z(2^{m+1})/2$; that is, 2^{m-2} in $\Phi_Z(2^m)$ and 2^{m-1} in $\Phi_Z(2^{m+1})$. Then the given isomorphisms establish the truth of the lemma. □

Now since $[-1]$ is not in K and since $[i]$ generates I, K intersects I only at the identity. Since each member of $K \times I$ can be represented by a real number or a pure imaginary number,

$H \cap (K \times I) = \{[1]\}$. The order of $H \times K \times I$ is, therefore,

$$2^{m-1+m-2+2} = 2^{n-1} = \phi_G(\alpha^n) \text{ if } n = 2m,$$
$$2^{m-1+m-1+2} = 2^{n-1} = \phi_G(\alpha^n) \text{ if } n = 2m+1.$$

Then $\Phi_G(\alpha^n) = H \times K \times I$, and we have proved the following theorem.

THEOREM 6. *Let $n \geqslant 5$. Then*

$$\Phi_G(\alpha^n) \simeq Z_{2^{m-1}} \times Z_{2^{m-2}} \times Z_4 \text{ if } n = 2m,$$
$$\Phi_G(\alpha^n) \simeq Z_{2^{m-1}} \times Z_{2^{m-1}} \times Z_4 \text{ if } n = 2m+1.$$

8. Φ_G Is Multiplicative. Question 3 of the Introduction has been answered by Theorems 3–6. We will show now that

$$\Phi_G(\beta_1\beta_2) \simeq \Phi_G(\beta_1) \times \Phi_G(\beta_2)$$

when β_1 is prime to β_2, thus answering question 4 in the affirmative. Now it is easy to modify a "Z" proof of a version of the Chinese Remainder Theorem so as to establish that if β_1, β_2, η_1, and η_2 are members of G with β_1 prime to β_2, then there exists X in G satisfying $X \equiv \eta_1 \pmod{\beta_1}$ and $X \equiv \eta_2 \pmod{\beta_2}$. Moreover, X is unique (mod $\beta_1\,\beta_2$). These facts are helpful in establishing the following theorem.

THEOREM 7. *Let β_1 and β_2 be in G with $(\beta_1, \beta_2) = 1$. Let f map $\Phi_G(\beta_1) \times \Phi_G(\beta_2)$ to $G/(\beta_1\beta_2)$ such that*

$$f(([\eta_1], [\eta_2])) = [\eta],$$

where $\eta \equiv \eta_i \pmod{\beta_i}$ for $i = 1$ or 2. Then the image of f is $\Phi_G(\beta_1\beta_2)$ and f is an isomorphism.

Proof. The Chinese Remainder Theorem implies that f is well defined. For $i = 1$ or 2, $(\eta, \beta_i) = (\eta_i, \beta_i) = 1$ since $\eta \equiv \eta_i \pmod{\beta_i}$ and η_i is prime to β_i, being a member of $\Phi_G(\beta_i)$. Thus η is prime to $\beta_1\beta_2$ and f maps to $\Phi_G(\beta_1\beta_2)$. A routine check verifies that f is an isomorphism onto $\Phi_G(\beta_1\beta_2)$. □

EXAMPLE 8. Assuming that it might be fun to see the function f in action, we apply f to $\Phi_G(3) \times \Phi_G(\alpha^3)$. We know that $\Phi_G(3)$ is cyclic with generator $[1+i]$. (See the Remark following Theorem 4.) From Theorem 5 we have that $\Phi_G(\alpha^3) = \Phi_G(2\alpha)$ is cyclic with generator $[i]$. In view of Examples 4 and 5 and Theorems 1 and 2, we might conjecture that $G/(6\alpha) =$

$$\{[a+bi] : 0 \leqslant a \leqslant 11 \text{ and } 0 \leqslant b \leqslant 5\}$$

and that $[a+bi]$ is a unit if $a \not\equiv b \pmod 2$ and $(3, a) = 1$ or $(3, b) = 1$. This conjecture does check out, and a systematic listing of a set of representatives of $\Phi_G(6\alpha)$ might go as follows: $i, 5i, 1, 1+2i, 1+4i, 2+i, \ldots, 11+4i$. This set has 32 members, checking that $\phi_G(6\alpha) = \phi_G(3)\phi_G(\alpha^3) = 8(4)$. Now using the notation of Theorem 7 with $\beta_1 = 3$ and $\beta_2 = 2\alpha$, we let

$$\eta = \eta_1 + (3+6i)(\eta_2 - \eta_1).$$

(One who works through a proof of the Chinese Remainder Theorem can see the motivation.) Then $\eta \equiv \eta_1 \pmod 3$ and $\eta \equiv \eta_2 \pmod{2\alpha}$. For example, to find $f(([1+i], [3]))$, we let $\eta_1 = 1+i$ and $\eta_2 = 3$. Then

$$\eta = 1 + i + (3-6i)(2-i) = 1 - 14i \equiv 7 + 4i \pmod{6\alpha},$$

since $12i \equiv 0 \pmod{6\alpha}$ and $7 + 4i = 1 - 2i + 6\alpha$. Thus $f(([1+i], [3]))$ is $[7+4i]$. Fig. 2 is a display of f in its entirety. In this display we omit the brackets that denote equivalence classes. In the upper left corner we isolate the subgroup $\Phi_G(3) \times \{[1]\}$, which is, of course, isomorphic to $\Phi_G(3)$; in the upper right we isolate $\{[1]\} \times \Phi_G(2\alpha)$. We denote the images of these subgroups by S and T. The reader who wishes to practice his computational skills is urged to show directly that

$[1 + 4i]$ in $\Phi_G(6\alpha)$ generates S and $[10 + 3i]$ generates T. It is clear that $S \cap T = \{[1]\}$, so that $\Phi_G(6\alpha) = S \times T$.

$$\Phi_G(3) \times \Phi_G(2\alpha) \xrightarrow{f} \Phi_G(6\alpha)$$

	$\xrightarrow{f}$	
$(1,1)$	1	S
$(1+i,1)$	$1+4i$	S
$(2i,1)$	$3+2i$	S
$(1+2i,1)$	$7+2i$	S
$(2,1)$	5	S
$(2+2i,1)$	$11+2i$	S
$(i,1)$	$9+4i$	S
$(2+i,1)$	$5+4i$	S
$(2,3)$	11	
$(2,i)$	$2+3i$	
$(2,2+i)$	$8+3i$	
$(i,3)$	$3+4i$	
(i,i)	i	
$(i,2+i)$	$6+i$	
$(2i,3)$	$9+2i$	
$(2i,i)$	$5i$	

	$\xrightarrow{f}$	
$(1,1)$	1	T
$(1,i)$	$10+3i$	T
$(1,3)$	7	T
$(1,2+i)$	$4+3i$	T
$(2i,2+i)$	$6+5i$	
$(1+i,3)$	$7+4i$	
$(1+i,i)$	$4+i$	
$(1+i,2+i)$	$10+i$	
$(1+2i,3)$	$1+2i$	
$(1+2i,i)$	$4+5i$	
$(1+2i,2+i)$	$10+5i$	
$(2+i,3)$	$11+4i$	
$(2+i,i)$	$8+i$	
$(2+i,2+i)$	$2+i$	
$(2+2i,3)$	$5+2i$	
$(2+2i,i)$	$8+5i$	
$(2+2i,2+i)$	$2+5i$	

FIG. 2.

9. Primitive Roots. Theorem 3 says that $\Phi_G(\pi^n)$ is cyclic, Theorem 4 implies that $\Phi_G(p^n)$ is cyclic only if $n = 1$, while Theorems 5 and 6 yield that $\Phi_G(\alpha^n)$ is cyclic only if $n \leqslant 3$. We summarize:

Group	Order	Cyclic?
$\Phi_G(\pi^n)$	$q^{n-1}(q-1)$	yes
$\Phi_G(p^n)$	$p^{2n-2}(p^2-1)$	only if $n = 1$
$\Phi_G(\alpha^n)$	2^{n-1}	only if $n \leqslant 3$

Thus α, α^2, α^3, π^n, and p have primitive roots. The product of cyclic groups is cyclic if and only if their orders are relatively prime. Then $\alpha\pi^n$ and αp also have primitive roots, but these are the only additional ones that we get in this manner. We therefore have the following theorem.

THEOREM 8. *The Gaussian integers α, α^2, α^3, π^n, p, $\alpha\pi^n$, and αp have primitive roots. These and their associates are the only Gaussian integers having primitive roots.*

As we mentioned in the Introduction, Crowe proves Theorem 8. Since his paper is unpublished, we comment on his methods. Versions of Theorems 1–3 appear in his work, so that he obtains that $\Phi_G(\pi^n)$ is cyclic. He then observes that $\Phi_Z(2^m)$ or $\Phi_Z(2^{m+1})$ is included isomorphically in $\Phi_G(\alpha^n)$, implying that α^n has no primitive root if $n > 3$. Finally, for $n > 1$, he finds two distinct subgroups of order p in $\Phi_G(p^n)$, implying that $\Phi_G(p^n)$ is not cyclic.

References

1. Ethan D. Bolker, Elementary Number Theory, W. A. Benjamin, New York, 1970.
2. Michael S. Crowe, An extension of the Euler ϕ-function to the Gaussian Integers, Senior Honors paper, 1974, unpublished. Available at the duPont Library, The University of the South.
3. Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 4th ed., Wiley, New York, 1980.
4. Henry B. Mann, Introduction to Algebraic Number Theory, The Ohio State University Press, 1955.

GENERIC GEOMETRY

J. W. BRUCE
Department of Mathematics, University College, Cork, Ireland

P. J. GIBLIN*
Department of Pure Mathematics, University of Liverpool, Liverpool L69 3BX, England

While it is impossible to say anything about everything,
one can say almost everything about almost anything.

Introduction. This article is about smooth curves embedded in the plane. We stick to this simplest (but far from trivial) case here, though our methods apply also to higher dimensions. An embedded curve C in the plane does not cross itself, and does not have kinks or cusps. (From Section 1, we only consider *closed* curves, without loose ends.) It does have a tangent line at each point. Some tangent lines have closer contact than others: the tangent line to $y = f(x)$ at $(x_0, f(x_0))$ is said to have *n-point contact* iff the ith derivative, $f^{(i)}(x_0)$, of f at x_0 is zero for each i with $2 \leqslant i \leqslant n-1$ and nonzero for $i = n$. *Ordinary* or *2-point* contact just means $f''(x_0) \neq 0$; higher contact is also called *inflexional* contact (Fig. 1). Thus, for example, when n is an integer $\geqslant 1$, the curve $y = x^n$ has exactly n-point contact with its tangent line (the x-axis) at $(0,0)$

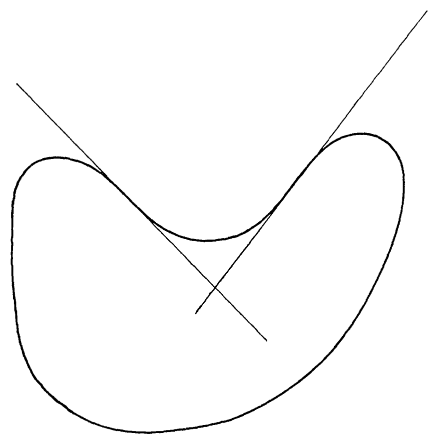


FIG. 1. Inflexional tangents.

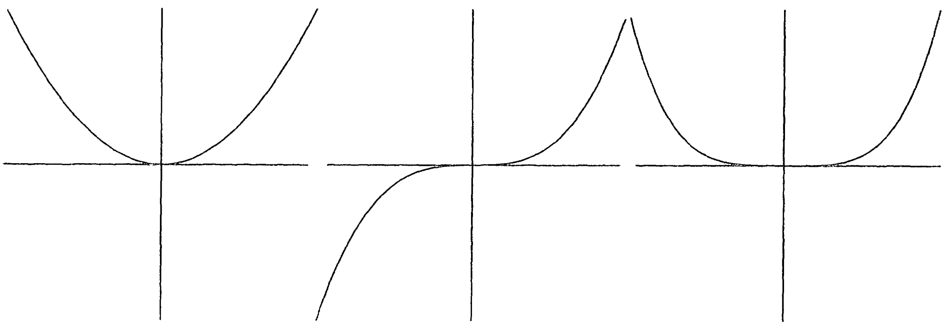
Biographical note prepared by Peter Giblin at the request of the editors:

Bill Bruce was born in Liverpool and took his Bachelor's, Master's, and Doctor's degrees at the University there. In between he spent a year teaching at a training college and a year schoolteaching, and after his doctorate was a year-long visitor at the Institut des Hautes Etudes Scientifiques, Paris, and then a postdoctoral fellow at Liverpool, working with the singularity theory group headed by C. T. C. Wall. He is interested in the applications of singularity theory to geometry; he and I collaborated with C. G. Gibson at Liverpool on an extensive investigation of the geometry of light caustics, starting with an article in this MONTHLY, November 1981. We have recently written a book developing geometrical applications of singularity theory for undergraduates.

I have been at Liverpool since 1967, where I moved after completing my Ph.D. at London University. I work with the singularity theory group at Liverpool, and during the year 1981–82 I was Visiting Professor at the University of North Carolina, Chapel Hill.

Bill plays on a badminton team for Killingley Club, Cork. I play the piano. We both enjoy computer graphics. (The pictures in the article were all done by me, with some help from a computer, at Liverpool.)

*The second author was a visitor at University of North Carolina, Chapel Hill, while this article was written.



$y = x^2$ $y = x^3$ $y = x^4$

FIG. 2.

(Fig. 2), and for $n \geqslant 3$ this contact is inflexional.

Sometimes a single line can be a *bitangent* of C (tangent at two different points of C), or even a *tritangent* (Fig. 3).

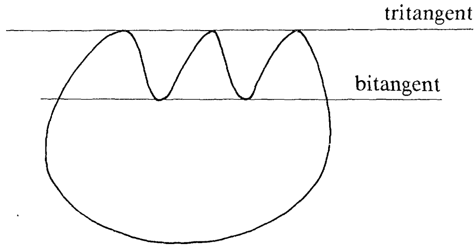


FIG. 3.

Here are two results about tangents to plane curves.

(a) Suppose that the tangents to C are *all* inflexional. Then C is a straight line (or several separate straight lines).

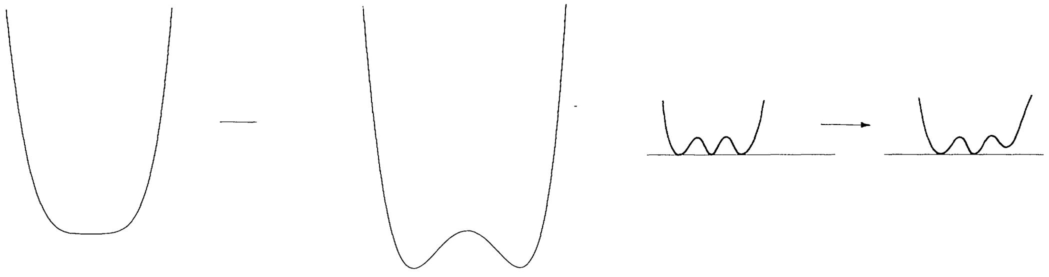
(b) Almost all plane curves possess *no* tangents whose order of contact is $\geqslant 4$ and *no* tritangents.

The results (a) and (b) could hardly be more different. The first has an extremely strong hypothesis, and tells us, not surprisingly, that virtually nothing satisfies this hypothesis—only straight lines. It is typical of the results of differential geometry of the form "if something (a curve) is sufficiently like something else (a line) at every point, then it *is* something else."

Result (b) cannot be accused of having a strong hypothesis—indeed, it may not be entirely clear whether it has one at all. It asserts that, "in general," plane curves do not have peculiarities such as 4-point contact tangents or tritangents. If by accident they do, then these peculiarities can be removed by a tiny change in the curve—they are *unstable* (Fig. 4)—whereas the properties in (b) are *stable*—not removable by small perturbations. Generic Geometry is the study of properties of smooth curves, surfaces and higher dimensional manifolds in a euclidean space which hold "in general." We shall be more precise in § 4.

Although 19th century geometers were interested in results which hold in general, the techniques to make these results precise and to prove them were not available until relatively recently. Following suggestions by René Thom, some very interesting results were proved by Ian Porteous [13] and Eduard Looijenga [11] in the early 1970's. Since then many mathematicians have contributed to this exciting and significant development in differential geometry.

We hope to convey the flavor of the subject in this article. In §§ 1–4, we take the reader



$y = x^4$ $\qquad$ $y = x^4 - \varepsilon x^2$

FIG. 4.

through a proof of part of result (b) which is more direct and geometrical than proofs obtained by specializing more general theorems in the literature. We use the crucial concept of transversality, introduced in § 2, and a device which we call the Taylor map, introduced in § 1. After that, in §§ 5–6, we interpret result (b) as a theorem about duals and evolutes, using another keystone of modern singularity theory, versal unfoldings, which have here a simple and striking application. We cannot include all the details, of course, but we hope that the main themes are clear.

We have no space for the more challenging results of Generic Geometry in higher dimensions, but recommend to the reader references [**2**], [3], [**5**], [**11**], [**13**], [**14**]. In fact [3], [**14**] are good general references. Many of the ideas sketched here are presented in more detail in [7]. A good introduction to singularity theory is [**4**], while the underlying theory of smooth manifolds is presented in [**12**], [**10**].

1. The Taylor Map. Throughout this article we consider smooth (infinitely differentiable) functions and curves.

Let $e\colon S^1 \to \mathbb{R}^2$ be a smooth embedding of the unit circle S^1 into the plane $\mathbb{R}^2$. (Thus using, say, the angle coordinate t on S^1 and writing $e(t) = (X(t), Y(t))$, the functions X and Y are smooth of period 2π, the derivatives $X'(t)$ and $Y'(t)$ are never both zero for the same t, and distinct values of t map to distinct points of $\mathbb{R}^2$.) We call e or its image $e(S^1) = C$ an *embedded curve* in the plane. These embedded curves are the object of our study.

Let V_k be the set of polynomials in a single variable (call it ξ, but for the most part it can remain anonymous) of the form

$$a_2\xi^2 + a_3\xi^3 + \cdots + a_k\xi^k$$

where the a_i are real numbers. Thus, V_k is a real $(k-1)$-dimensional vector space, which we can think of as $\mathbb{R}^{k-1}$ with coordinates $(a_2, a_3, \ldots, a_k)$.

One of our main tools will be what we call the *Taylor map*

$$\gamma\colon C \to V_k$$

defined as follows. Choose a point $p = e(t)$ of C. At p there is a unit tangent vector $T(t)$ in the direction $(X'(t), Y'(t))$ and a unit normal vector $N(t)$ in the direction $(-Y'(t), X'(t))$. This gives us two axes at p, in the T and N directions, and we can write the equation of C, close to p at any rate, as, say, $\eta = f(\xi)$ relative to those axes. (Thus, ξ is along the T direction, η along N.) See Fig. 5. Since the ξ-axis is tangent to C at p, we have $f(0) = f'(0) = 0$ where $'$ here means $d/d\xi$. Of course, f depends on p, but for a given p, f is uniquely determined. (Clearly it is also unchanged if we reparametrize the curve C.) We can form the Taylor series (Maclaurin series) of f at $\xi = 0$, carried as far as the term ξ^k. This gives us the polynomial $\gamma(p)$ which we seek.

As a simple example, if C is a circle of radius r, then for $k = 3$, $a_2 = 1/2r$ and a_3 is zero, no matter what p is. Thus, γ has image consisting of the single point $(1/2r, 0) \in V_3$ in this case. Other examples are rather harder to work out explicitly! See Fig. 6 below, and (2.2).

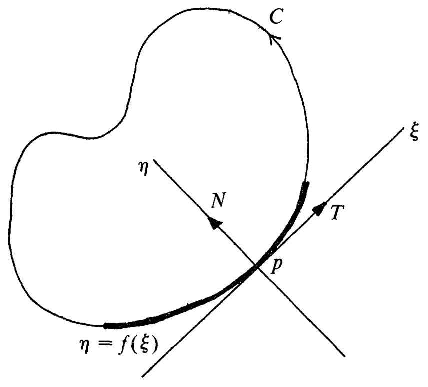


FIG. 5. Setting up the Taylor map of C.

The coefficients (or coordinates) a_i have interpretations in terms of the *curvature* of C at p, and its derivatives. (See, for example, [**9**].) In fact, using the local equation $\eta = f(\xi)$, i.e., the parametrization $\xi \mapsto (\xi, f(\xi))$, the curvature at $(\xi, f(\xi))$ of C is

$$\kappa(\xi) = f''(\xi)/\left(1 + (f'(\xi))^2\right)^{3/2}.$$

It follows quite easily that $\kappa(0) = 2a_2$, $\kappa'(0) = 6a_3$ and $\kappa''(0) = 24(a_4 - a_2^3)$. (Remember that $'$ means $d/d\xi$. Actually, for the first two derivatives of κ, the derivative with respect to arc-length on C gives the same answers.) Thus, p is an *inflexion* of C iff $a_2 = 0$; a *higher* (or *degenerate*) *inflexion*, where the tangent line has at least 4-point contact, iff $a_2 = a_3 = 0$; a *vertex* of C (point of stationary curvature) iff $a_3 = 0$; and a *higher* (or *degenerate*) *vertex* iff $a_3 = a_4 - a_2^3 = 0$. An inflexion or vertex which is *not* higher is called *ordinary*.

Let us take $k = 3$ and consider our map γ: $C \to V_3$. We picture V_3 as a plane with coordinates

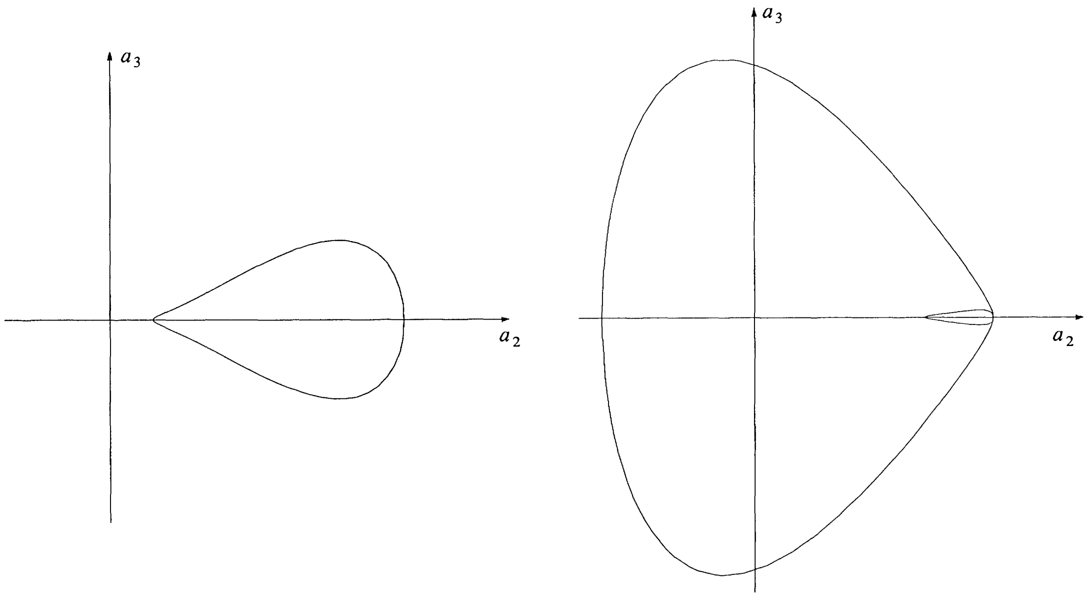


Image of Taylor map of ellipse. Image of Taylor map of limaçon.

FIG. 6.

(a_2, a_3). The image of γ will be a curve $\gamma(C)$ in V_3—or it might collapse as in the case of C a circle. If $\gamma(C)$ passes through the origin, then C has a higher inflexion; if $\gamma(C)$ crosses the a_2-axis, then C has a vertex, and so on. Fig. 6 shows the curve $\gamma(C)$ for an ellipse and for a limaçon. (For the ellipse, $\gamma(C)$ is traversed *twice*: there are four vertices.)

Suppose $\gamma(C)$ does pass through the origin. Then, we would expect that by slightly deforming C to C', say, $\gamma(C')$ could be made to miss the origin: it should be "accidental" that $\gamma(C)$ passes through the selected point $(0,0)$. Thus, we expect that curves with higher inflexions will be exceptional.

Likewise with $k = 4$, using $\gamma\colon C \to V_4$, we obtain a curve $\gamma(C)$ in the 3-space with coordinates (a_2, a_3, a_4). (Again $\gamma(C)$ might collapse, of course.) The set with equations $a_3 = a_4 - a_2^3 = 0$ in V_4 is a curve (see Fig. 7) and "in general" we would expect $\gamma(C)$ to miss this curve: if $\gamma(C)$ does happen to intersect it, then we would expect that by changing C slightly to C', we could arrange for $\gamma(C')$ to miss the curve. Thus, we would expect a "general" C to have no higher vertices. The set $a_3 = 0$ is a plane in V_4, and if $\gamma(C)$ meets this plane (but does not touch it), we would expect nearby curves $\gamma(C')$ to continue to meet the plane. (See Fig. 8.) This suggests that ordinary vertices persist under small changes in C.

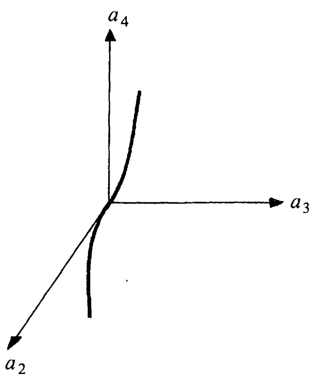


FIG. 7. The higher vertex set $a_3 = a_4 - a_2^3 = 0$.

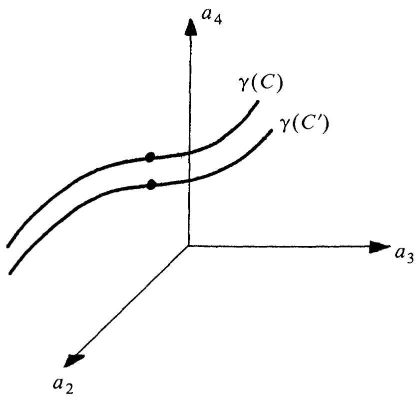


FIG. 8.

To sum up, it seems reasonable that the following properties will hold "in general" for curves C in the plane:

(1.1) (i) C has no higher inflexions,
(ii) C has no higher vertices.

It is pretty clear that there are objections to this kind of argument: apart from its general air of vagueness, it is far from obvious how changes in C are going to affect the curve $\gamma(C)$. There are indeed hidden constraints on the curve $\gamma(C)$: for instance, when $k = 3$, $\gamma(C)$ cannot touch the axis $a_2 = 0$ except at the origin, and a famous theorem called the *four vertex theorem* [**9**, p. 36] implies that at least four points of C must map to the axis $a_3 = 0$.

In the following sections, we try to show how modern singularity theory overcomes these obscurities. We hope that an occasional lack of detail in what follows (necessitated by pressure of space) will not be mistaken for a more complicated form of obscurity.

2. Transversality. Two embedded curves C_1 and C_2 in the plane are said to be *transverse* at an intersection point q if their tangent lines at q are different. This is the same as saying that the tangent vectors T_1 and T_2 to the curves at q span all the "tangents to the plane" at q: every vector based at q is a linear combination $\alpha T_1 + \beta T_2$ of T_1 and T_2 with real coefficients α, β. See Fig. 9.

More generally in a Euclidean space $V = \mathbb{R}^2$ and $\mathbb{R}^3$ or V_k (regarded as $\mathbb{R}^{k-1}$), two smooth *submanifolds* Q_1 and Q_2, (think of each as a point, a curve or a surface) are *transverse* at an

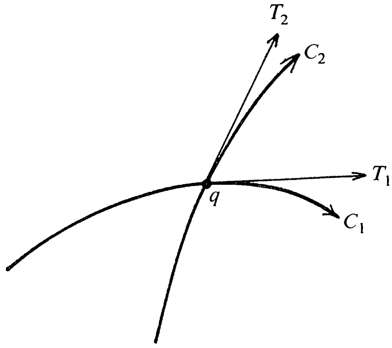


FIG. 9. Transverse curves.

intersection point $q \in V$ provided every vector based at q is a linear combination of a tangent vector to Q_1 and a tangent vector to Q_2. If Q_1 is a single point q, then it has no nonzero tangent vectors, so when $q \in Q_2$, Q_2 is transverse to q provided every vector based at q is tangent to Q_2. In that case Q_2 contains a neighborhood of q.

We call Q_1 and Q_2 *transverse* if they are transverse at every $q \in Q_1 \cap Q_2$. This has the usual mathematical quirk that if Q_1 and Q_2 fail to intersect, then they are automatically transverse, for there cannot possibly be a q at which they fail to be transverse!

Two surfaces in $\mathbb{R}^3$, or a curve and a surface in $\mathbb{R}^3$ are transverse at a common point q provided they do not just touch at q: one cuts cleanly through the other. Two curves in $\mathbb{R}^3$ can never be transverse at a common point, for the dimensional reason that two vectors are not enough to span a three-dimensional space. Thus if we can prove that two curves in $\mathbb{R}^3$ *are* transverse, then the curves *cannot intersect*. The same goes for a curve and a point in $\mathbb{R}^2$ (or $\mathbb{R}^3$): if they are transverse, then *the point cannot lie on the curve*.

For the Taylor map $\gamma\colon C \to V_k$ as in §1, we want to say that γ is transverse to a submanifold (point, curve, surface...) Q in V_k if Q and $\gamma(C)$ are transverse. Because C may get collapsed by γ, we have to spell this out more carefully.

Let $p_0 \in C$ have parameter say $t_0 \in S^1$ so that $p_0 = e(t_0)$. Then as the real variable s moves away from zero, $t_0 + s$ gives us a *path* in C, consisting of points $p_s = e(t_0 + s)$. Using γ, we get a path in V_k consisting of points $\gamma(p_s)$. The tangent vector to C at p_0 is taken, by definition, to the vector

$$v = \lim_{s \to 0} \frac{\gamma(p_s) - \gamma(p_0)}{s} \tag{2.1}$$

based at $\gamma(p_0)$ in V_k. (Thus v is the derivative of $\gamma \circ e$ at t_0.) See Fig. 10.

If $\gamma(C)$ is a genuine embedded curve in V_k, then the vector v is simply its tangent vector at

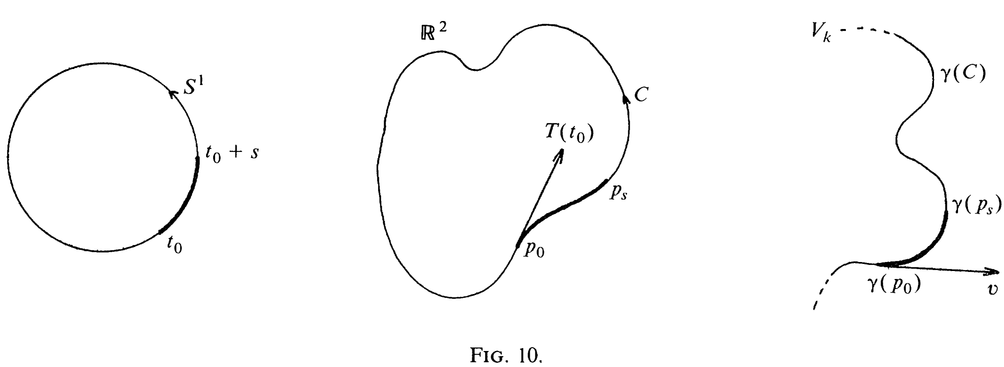


FIG. 10.

$\gamma(p_0)$, but the definition works in any case. For example, if C is a circle, then v is always zero.

We say that γ *is transverse to* Q provided the following holds for every $p_0 \in C$ with $\gamma(p_0) = q_0 \in Q$: every vector based at q_0 in V_k is a linear combination of the above vector v and tangent vectors to Q at q_0.

(2.2) *Example.* Suppose that C is given, close to $(0,0)$, by the equation $y = x^2$, which we can parametrize by $e(t) = (t, t^2)$. Take $p_0 = (0,0)$, $p_s = (s, s^2)$. Let $k = 3$ and take for Q the a_2-axis in V_3, which has tangent vector $(1,0)$ (along the a_2-axis) at $q_0 = (1,0) \in V_3$. (See Fig. 11.) A straightforward calculation shows that

$$\gamma(p_s) = \left((1 + 4s^2)^{-3/2}, -4s(1 + 4s^2)^3\right)$$

so that the vector v defined in (2.1) works out at $(0, -4)$. Since every vector in $\mathbb{R}^2$ is a linear combination of $(1,0)$ and $(0, -4)$, γ is indeed transverse to the a_2-axis at q_0.

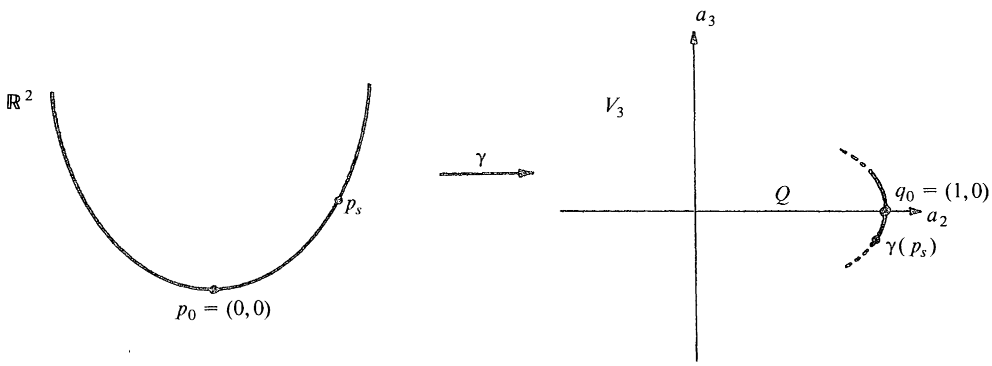


FIG. 11. Example (2.2).

One of our objectives is to show that, for "most" curves C, γ is transverse to the origin for $k = 3$ and to the curve with equations $a_3 = a_4 - a_2^3 = 0$ for $k = 4$. From these and the definition of transversality it will follow that $\gamma(C)$ does not meet these sets at all, and this implies (i) and (ii) of (1.1). Actually we shall prove a much more general transversality statement.

3. Families of Curves. Starting with an embedded curve C in $\mathbb{R}^2$, we can deform C by means of a collection U of smooth maps $\mathbb{R}^2 \to \mathbb{R}^2$. Thus if $\psi \in U$ and C is given by $e\colon S^1 \to \mathbb{R}^2$, then $\psi(C)$ is given by $\psi \circ e$, which takes $t \in S^1$ to $\psi(e(t)) \in \mathbb{R}^2$. We shall always take U to be an open set of some euclidean space $\mathbb{R}^n$. For example, U could consist of all translations of $\mathbb{R}^2$

$$\left(\psi(x, y) = (x + c_1, x + c_2), U = \mathbb{R}^2 = \{(c_1, c_2)\}\right)$$

or all affine maps

$$\left(\psi(x, y) = (c_1 + c_2 x + c_3 y, c_4 + c_5 x + c_6 y), U = \mathbb{R}^6 = \{(c_1, \dots, c_6)\}\right).$$

Suppose ψ is such that $\psi(C)$ is still an embedded curve. Then there is a corresponding map

$$\gamma = \gamma_\psi\colon \psi(C) \to V_k.$$

All of these can be fitted into a family

$$\Gamma\colon C \times U \to V_k$$

defined by $\Gamma(p, \psi) = \gamma_\psi(\psi(p))$. In other words, given $p \in C$ and a transformation ψ we work out the Taylor expansion of the transformed curve $\psi(C)$ at the point $\psi(p)$ and write down its terms up to degree k (or just the coefficients of those terms).

As an example, let C be the circle given by $e(t) = (\cos t, \sin t)$, and let $\psi(x, y) = (x, (\lambda + 1)y)$ where $\lambda \in \mathbb{R}, \lambda \neq -1$. Then ψ takes C to the ellipse $\psi(C)$ consisting of points $(\cos t, (\lambda + 1)\sin t)$.

As λ moves away from zero, the image $\gamma_\psi(\psi(C))$ in V_3 changes from a point $(\frac{1}{2}, 0)$ to a genuine curve, as illustrated in Fig. 12.

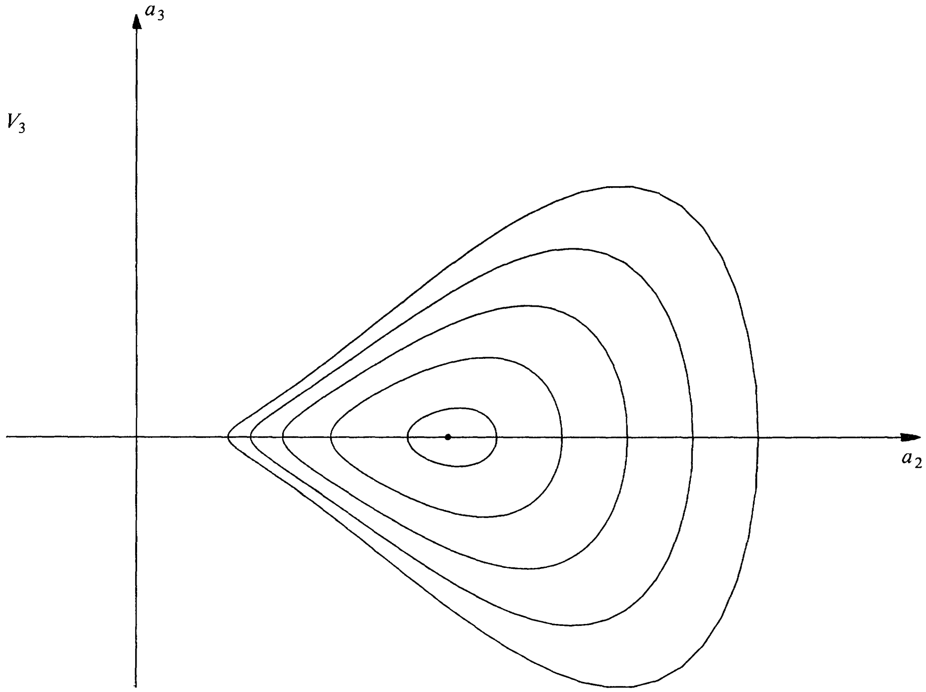


FIG. 12. Images of the Taylor map for a family of ellipses.

Just as we can speak of $\gamma\colon C \to V_k$ being transverse to a submanifold Q of V_k, we can speak of the whole family Γ being transverse to Q. Here is a simple example.

(3.1) Let C be as in (2.2). Let U be the set of maps $\mathbb{R}^2 \to \mathbb{R}^2$ of the form $\psi_\lambda(x, y) = (x, (\lambda + 1)y)$ where λ is so close to 0 that $\psi_\lambda(C)$ is still an embedded curve. Let us examine the map

$$\Gamma\colon C \times U \to V_3$$

which takes (p, ψ_λ) to the coefficients (a_2, a_3) in the Taylor expansion of $\psi_\lambda(C)$ at $\psi_\lambda(p)$. Note that $\Gamma(p_0, \psi_0) = (1,0) = q_0$, say. We shall suppose for simplicity that no other (p, ψ) goes by Γ to q_0. The image of Γ is a whole system of curves in V_3 parametrized by λ. Taking the path $s \to (p_0, \psi_s)$ in $C \times U$, we obtain from Γ the path $s \to \Gamma(p_0, \psi_s) = (s + 1, 0)$ in V_3. See Fig. 13. (For $\psi_s(C)$ is the curve $y = (s + 1)x^2$ close to $(0,0)$, and at $\psi_s(p_0) = p_0$ the Taylor expansion $(s + 1)x^2 + 0x^3$ gives $a_2 = s + 1$, $a_3 = 0$.) Hence, the path in V_3 gives us the vector

$$\lim_{s \to 0} \frac{(s + 1, 0) - (1, 0)}{s} = (1, 0)$$

in V_3 based at q_0. This is the image under Γ of a *tangent vector* to $C \times U$ at (p_0, ψ_0). We already know from (2.2) that the path $s \to (p_s, \psi_0)$ in $C \times U$ gives us the vector $(0, -4)$ at q_0. Because $(1,0)$ and $(0, -4)$ span all vectors in $\mathbb{R}^2$ this shows that the whole family Γ is transverse to the point q_0 in V_3.

This is the kind of result we shall need. The reason is the following crucial lemma. This lemma does not assert something about *all* members of the set U which parametrizes our family of curves, but about *almost all* members. This phrase has a precise technical meaning, which we say something about below, but the reader will have the right idea if he thinks of "all but a tiny minority of elements of U."

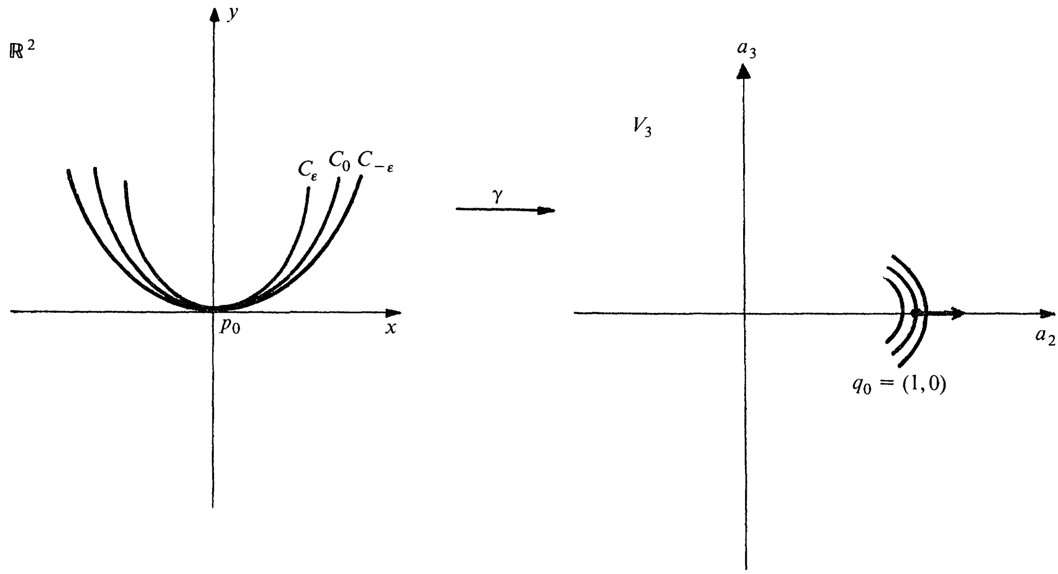


FIG. 13. Example of a family of curves.

(3.2) TRANSVERSALITY LEMMA OF SARD AND THOM. *Suppose that* Γ: $C \times U \to V$ *is transverse to the submanifold* Q *of* V. *Then for almost all* $\psi \in U$ *the map* Γ_ψ: $C \to V$ *defined by* $\Gamma_\psi(p) = \Gamma(p, \psi)$ *is transverse to* Q.

Note that $\Gamma_\psi(p) = \gamma_\psi(\psi(p))$ where γ_ψ is the Taylor map of the curve $\psi(C)$.

By "almost all" we mean all $\psi \in U$ except those of a *set of measure zero in* U. The precise technical meaning of this standard term from the theory of Lebesgue measure need not concern us. A finite or countable set of points in a line has measure zero in the line; a finite or countable set of points and lines has measure zero in the plane, as has any smooth curve C. We do need to know that if Ω has measure zero in U, then $U - \Omega$ is *dense* in U: every point of U has points of $U - \Omega$ arbitrarily close to it. Thus, a "bad" $\psi \in \Omega$ is arbitrarily close to "good" $\psi' \in U - \Omega$. We also need to know that if $\Omega_1, \dots, \Omega_l$ have measure zero, then so does their union: almost all ψ fail to belong to *any* of the Ω_i.

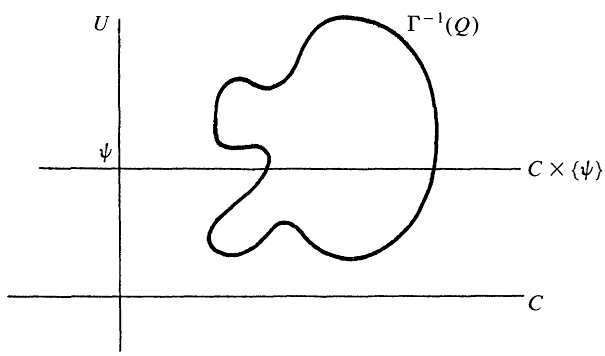


FIG. 14. Pictorial proof of (3.2).

Here is a pictorial indication of why (3.2) is true—actually the formal proof ([**14**], p. 739) is surprisingly similar. The fact that Γ is transverse to Q actually makes $\Gamma^{-1}(Q)$ a smooth manifold in $C \times U$ (see Fig. 14). It turns out that Γ_ψ is transverse to Q if and only if $C \times \{\psi\}$ intersects $\Gamma^{-1}(Q)$ transversely. As is plausible from the picture, *most* ψ will give a "horizontal line" $C \times \{\psi\}$

which is transverse to, i.e., fails to touch, $\Gamma^{-1}(Q)$. This last is actually a consequence of Sard's theorem (see, for example, [12]) on critical values. Thom's crucial idea was to turn Sard's theorem into a transversality statement.

In the example (3.1) above, with $Q = \{q_0\}$, the only map Γ_ψ whose image passes through Q at all is with $\psi = \psi_0$. Now Γ_{ψ_0} is *not* transverse to Q but every other Γ_ψ *is* transverse to Q (because there is no point where transversality has to be checked). Thus, the "bad" ψ make up one point $\lambda = 0$ out of a whole interval of values of λ, and indeed a point has measure zero in an interval.

If we want transversality to several different submanifolds $Q_1, \ldots, Q_l$ say, then we can apply Thom's lemma to each of the Q_i separately, obtaining l sets of measure zero in U. The union Ω of these l sets will still be of measure zero and for $\psi \in U - \Omega$ we shall have Γ_ψ transverse to all the Q_i. Thus, in order to show that, for almost all ψ, Γ_ψ is transverse to all the Q_i, we need only prove Γ is transverse to all the Q_i.

Let us return to the case of a single Q. The way in which one proves that Γ is transverse to Q is to take paths in $C \times U$ through $(p_0, \psi_0) \in \Gamma^{-1}(Q)$, as in (3.1), and to calculate the corresponding vectors at $\Gamma(p_0, \psi_0) \in V_k$. In fact, we shall always find, in the important examples below, that it is enough to take paths of the form $s \to (p_0, \psi_s)$ where only ψ_s varies.

4. Generic Properties of Curves. Let C be a curve embedded in $\mathbb{R}^2$ and let P_k be the set of maps $\psi\colon \mathbb{R}^2 \to \mathbb{R}^2$ of the form

$$\psi(x, y) = (\psi_1(x, y), \psi_2(x, y))$$

where each ψ_i is a polynomial in x and y whose degree is $\leqslant k$. Thus an element $\psi \in P_k$ is determined by the coefficients of the various monomials $x^i y^j$ occurring in ψ_1 and ψ_2. There are altogether $1 + 2 + \cdots + (k+1) = \frac{1}{2}(k+1)(k+2)$ monomials of degree $\leqslant k$ so P_k can be thought of as a euclidean space $\mathbb{R}^N$ where $N = (k+1)(k+2)$. Two elements ψ, ψ' of P_k are close together provided the coefficients of ψ_1 are close to the corresponding coefficients of ψ_1' and likewise for ψ_2, ψ_2'.

The identity map id: $\mathbb{R}^2 \to \mathbb{R}^2$, $\mathrm{id}(x, y) = (x, y)$ is of course an element of P_k provided $k \geqslant 1$ (we shall always have $k \geqslant 2$ in fact). Let U_1 consist of those $\psi \in P_k$ which are so close to the identity that $\psi(C)$ is still an embedded curve in $\mathbb{R}^2$. (Using the *compactness* of C (C is closed and bounded in $\mathbb{R}^2$) it actually follows that all elements of P_k sufficiently near id will be in U_1.) Thus,

$$\psi_1(x, y) = x + \phi_1(x, y), \qquad \psi_2(x, y) = y + \phi_2(x, y)$$

say, where all the coefficients in ϕ_1 and ϕ_2 are sufficiently small.

Let Q be a submanifold of V_k. We propose to prove:

(4.1) *For some open set U of U_1, containing the identity map, $\Gamma\colon C \times U \to V_k$ (see* §3) *is transverse to Q.*

In fact, Q does not enter the argument at all, for we show that we can obtain every vector based at $q = \Gamma(p, \psi)$ simply from tangent vectors to $C \times U$; tangent vectors to Q are not needed. (Γ is then called a *submersion*.)

First, a technicality. What we actually do is to take points (p, id) and show that every vector at $\Gamma(p, \mathrm{id})$ can be obtained from tangent vectors to $C \times U_1$. Using the compactness of C, a standard technical argument then shows that the same holds for (p, ψ) for all ψ in a sufficiently small neighborhood of the identity. (Alternatively a variant of the argument given shows that the same holds for any (p, ψ) with $\psi \in U_1$, which is better still.) The idea behind the proof of (4.1) is best illustrated by suppressing this part of the argument and sticking to points (p, id).

By choosing our coordinates, we can in fact assume that $p = (0,0) \in \mathbb{R}^2$ and that close to p the curve C is given by the equation $y = f(x)$, where $f(0) = f'(0) = 0$. Thus, $\Gamma(p, \mathrm{id})$ = terms of degree $\leqslant k$ in the Taylor series of f at 0. (To see this, let us do a rigid motion θ of $\mathbb{R}^2$ to bring C into standard position, with p moved to $(0,0)$ and the tangent line moved to the x-axis. This will

not affect the result because referring our transformations ψ to the new coordinates leaves them of the same degree, and also the Taylor expansion of C at p is the same as that of $\theta(C)$ at $\theta(p)$. The result about $\Gamma(p,\mathrm{id})$ now follows from the definition of Γ.)

Let $g(x)$ be a polynomial in x of degree $\geqslant 2$ and $\leqslant k$, and let $\psi_s(x, y) = (x, y + sg(x))$ where $s \in \mathbb{R}$. Then, for all small s, $\psi_s \in U_1$ and we can use the path $s \to (p, \psi_s)$ in $C \times U_1$. Now $\psi_s(t, f(t)) = (t, f(t) + sg(t))$ so $\psi_s(C)$ is still in the form where we can read off the Taylor expansion—this is why we chose g to have degree $\geqslant 2$. Namely, $\Gamma(p, \psi_s) =$ terms of degree $\leqslant k$ in the Taylor expansion of $f + sg$. The corresponding vector in V_k based at $\Gamma(p,\mathrm{id})$ is

$$\begin{aligned}\lim_{s\to 0} & \frac{(\text{terms of degree} \leqslant k \text{ in } f + sg) - (\text{terms of degree} \leqslant k \text{ in } f)}{s}\\ &= \text{terms of degree} \leqslant k \text{ in } g\\ &= g.\end{aligned}$$

By choosing g, we can therefore obtain every vector at $\Gamma(p,\mathrm{id})$ from paths in $C \times U$ and therefore the result is proved. □

Now we apply (4.1) and the Sard-Thom lemma (3.2) to some examples. (For some harder examples, see [**6**].)

(4.2) Take $k = 3$ and $Q = \{(0,0)\} \subset V_3$. Thus, the Taylor map $\gamma\colon C \to V_3$ takes C to a curve through Q if and only if C has a higher inflexion (compare § 1). Put γ in a family Γ as in (4.1), Γ being transverse to Q. Applying (3.2), we have that for a dense set of $\psi \in U$, $\Gamma_\psi\colon C \to V_3$ is transverse to Q, and hence $\Gamma_\psi(C)$ *misses* Q: for dimensional reasons transversality is impossible if they intersect. Hence, there are arbitrarily small deformations of C, given by such ψ arbitrarily close to the identity, which have *no* higher inflexions. The same result could be obtained by taking $k = 4$ and the *line* $a_2 = a_3 = 0$ as Q.

(4.3) Take $k = 4$ and let $Q = \{(a_2, a_3, a_4) \in V_4\colon a_3 = a_4 - a_2^3 = 0\}$. Thus, the Taylor map γ: $C \to V_4$ has $\gamma(C) \cap Q$ nonempty if and only if C has a higher vertex. Reasoning as before, we find that there are arbitrarily small deformations of C which have *no* higher vertices.

(4.4) As remarked in § 3, we can work with several submanifolds Q_i at once. Taking $k = 4$,

$$Q_1 = \{(a_2, a_3, a_4)\colon a_2 = a_3 = 0\}$$

and

$$Q_2 = \{(a_2, a_3, a_4)\colon a_3 = a_4 - a_2^3 = 0\},$$

we know that the Γ of (4.1) will be transverse to *both* Q_i, so applying (3.2) shows that there are arbitrarily small deformations of C which have neither higher inflexions nor higher vertices.

(4.5) This is marginally more technical. Take $k = 3$ and let Q_1 be the a_3-axis, Q_2 the a_2-axis. Then there exist ψ arbitrarily close to the identity for which $\gamma_\psi\colon \psi(C) \to V_3$ is transverse to Q_1 and Q_2. We can deduce from that (once again using the compactness of C) that $\gamma_\psi^{-1}(Q_1)$ and $\gamma_\psi^{-1}(Q_2)$ are *finite sets*. Hence, $\psi(C)$ *will have only a finite number of ordinary inflexions and ordinary vertices* ($a_2 = 0$ gives an inflexion and $a_3 = 0$ a vertex). There is no upper limit to the number of inflexions or vertices, of course. The 4-vertex theorem ([**9**], p. 36) asserts that there will be at least 4 vertices, but there could be no inflexions, as with a convex curve.

When we suggested in (1.1) that certain properties were "generic" for plane curves, we had in mind a specific meaning for "generic." Consider some property P of embedded curves e: $S^1 \to \mathbb{R}^2$, such as "possessing no higher inflexion" or "having a finite number of vertices." Generic geometers are interested only in geometrical properties, so we shall always take a P which is invariant under changes of parametrization and under rigid motions in $\mathbb{R}^2$.

(4.6) The property P is said to be *dense* (or, if you prefer, curves with property P are said to be dense in "the set of all curves") if every non-P curve can be taken by an arbitrarily small

deformation into a P-curve. More precisely, for each curve e, we can find a family

$$E: S^1 \times U \to \mathbb{R}^2$$

where U is a neighborhood of a point u_0 in some euclidean space (thought of as $\mathbb{R}^N$ for some N), such that

(i) $E(t, u_0) = e(t)$ for all $t \in S^1$, and

(ii) every neighborhood of u_0 in U contains a point u for which the curve given by E_u: $S^1 \to \mathbb{R}^2$ ($E_u(t) = E(t, u)$) has property P. (In the definition, the deformation does not *have* to be effected by a transformation $\mathbb{R}^2 \to \mathbb{R}^2$, but in all our examples that is how it is done.)

Thus, we have shown above that the properties of having no higher inflexions, no higher vertices (also finitely many inflexions or vertices) are dense. The space U consisted of polynomial transformations ψ, and $E(t, \psi) = \psi(e(t))$.

(4.7) The property P is said to be *open* if all curves close to a P-curve in a family are also P-curves. More precisely, for any embedded curve C, given by e as above, having the property P, and any family

$$E: S^1 \times U \to \mathbb{R}^2$$

with U as before and $E(t, u_0) = e(t)$ for all $t \in C$, the curves given by E_u for all u sufficiently close to u_0 also have property P.

For the properties we are considering, which can be described by transversality to certain closed submanifolds Q of V_k, openness is actually automatic, using (again) the compactness of C. Roughly speaking, if γ_e is transverse to Q, so is $\gamma_{e'}$ for any e' close to e. In fact, all the properties considered here are "open" in the slightly stronger sense that if C has one of these properties, then so does any curve given by an embedding "sufficiently close to e." We shall not go into this more technical matter here.

(4.8) A property P is called *generic*, or is said to hold *generically*, iff P is both open and dense. Note that if $P_1, \ldots, P_n$ are all generic, then so is the property $P = (P_1$ and P_2 and... and $P_n)$.

(4.9) *The following properties of plane curves are all generic*:

(i) *there are no higher inflexions*;

(ii) *there are no higher vertices*;

(iii) *there are finitely many* (*or no*) *ordinary inflexions*;

(iv) *there are finitely many* (*or no*) *ordinary vertices*. □

Notice that it is specific properties, and not specific curves, which are generic. A given curve may have some generic properties and some nongeneric ones (a circle has property (iii) above, but is highly nongeneric in its possession of vertices!). Also the list of generic properties can be extended indefinitely—and no curve could be so general as to possess *all* these properties! (Can you see why?) The important thing is that curves which have a specific property, known to be generic, constitute "most" curves in the open and dense sense explained above.

5. Duals and Evolutes. We now introduce some ideas from singularity theory to show how it is possible to deduce, from the genericity result (4.9), properties of duals and evolutes of plane curves. We give more detail for the former.

Let e: $S^1 \to \mathbb{R}^2$ give an embedded curve $C = e(S^1)$. We consider a *family of functions* on C parametrized by another space W, that is, we consider a map F: $S^1 \times W \to \mathbb{R}$.

(5.1) *Height functions*. Let W be the set of *directions* in $\mathbb{R}^2$, which we think of as given by unit vectors, i.e., $W = S^1$ so our family is of the form

$$H: S^1 \times S^1 \to \mathbb{R}.$$

Let $H(t, w) = e(t) \cdot w =$ "height" of $e(t)$ above the line through (0,0) perpendicular to w. See Fig. 15.

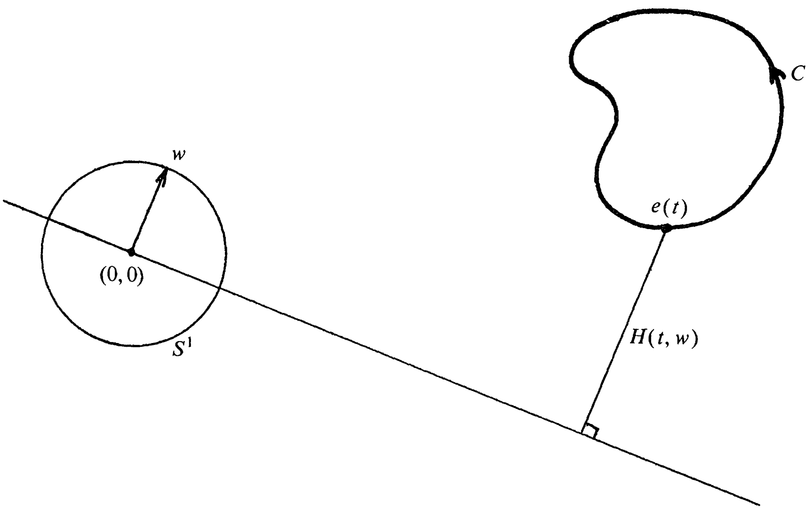


FIG. 15. Height function.

Consider the condition $\partial^i H/\partial t^i(t_0, w_0) = 0$ for $i = 1, \ldots, m$; $\neq 0$ for $i = m + 1$. When this holds we say that the function $h(t) = H(t, w_0)$, for the fixed w_0, has "type A_m at t_0."

To see what this means take $w_0 = (0, 1)$ ("vertical") and $e(t) = (t, f(t))$ so that C has equation $y = f(x)$. Then $H(t, w_0) = f(t)$ so the above condition just says that the tangent to C at $(t_0, f(t_0))$ is horizontal (parallel to the x-axis) and has $(m + 1)$-point contact with C. The same goes in general and we have, writing $N(t_0)$ for the normal to e at $e(t_0)$,

(5.2) $m = 1$ (type A_1) $\Leftrightarrow$ w_0 parallel to $N(t_0)$ but $e(t_0)$ not an inflexion of C
$m = 2$ (type A_2) $\Leftrightarrow$ " and $e(t_0)$ an *ordinary inflexion* of C
$m \geqslant 3$ (type $A_{\geqslant 3}$) $\Leftrightarrow$ " and $e(t_0)$ a *higher inflexion* of C. □

The set

$$CV(H) = \left\{(r, w) \in \mathbb{R} \times S^1 \colon \text{for some } t, \frac{\partial H}{\partial t}(t, w) = 0 \text{ and } H(t, w) = r\right\}$$

has an interesting interpretation. ($CV(H)$ is also the set of *critical values* of the map $S^1 \times S^1 \to \mathbb{R} \times S^1$ defined by $(t, w) \mapsto (H(t, w), w)$.) Let $T(t)$ be the unit tangent vector to the curve C at $e(t)$, so that $T(t)$ is a nonzero multiple of the vector de/dt.

Then $(\partial H/\partial t)(t, w) = 0 \Leftrightarrow T(t) \cdot w = 0 \Leftrightarrow w = \pm N(t)$ where $N(t)$ is the unit *normal* vector to C at $e(t)$, defined by turning $T(t)$ anticlockwise through 90°.

Oriented (directed) lines in $\mathbb{R}^2$ are in one-to-one correspondence with points $(r, w) \in \mathbb{R} \times S^1$, as follows. Given a line, let v be a unit vector parallel to the line, and define w by rotating v clockwise through 90°. Define $r = x \cdot w$ where x is any point of the line (Fig. 16). With this interpretation, $CV(H)$ falls into two parts: one consists of the tangent lines to C oriented via the embedding e (so that the tangent line at $e(t)$ has direction de/dt) and the other consists of the same lines oriented the other way (direction $-de/dt$). Choosing one of the two parts (say the first) we get the *dual* of C.

The dual thus lies in $\mathbb{R} \times S^1$ which is a cylinder. But locally we can flatten the cylinder on the plane, i.e., use the angle coordinate θ on S^1 so that a neighborhood of (r_0, w_0) in $\mathbb{R} \times S^1$ becomes a neighborhood of (r_0, θ_0) in $\mathbb{R} \times \mathbb{R}$, where $w_0 = (\cos\theta_0, \sin\theta_0)$.

We can also get a global picture of most of $CV(H)$ in $\mathbb{R}^2$ by projecting $\mathbb{R} \times S^1 \to \mathbb{R}^2$, $(r, w) \mapsto rw$ (this is like using polar coordinates). Points $(0, w) \in CV(H)$, which are occasioned by tangents to C passing through $(0,0)$, now all go to the origin, so the projection is not an

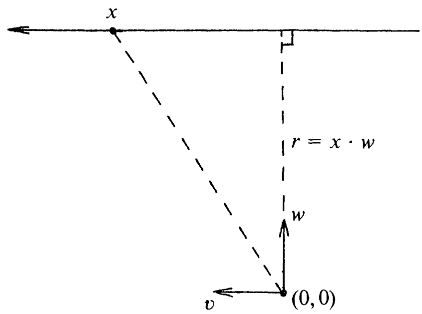
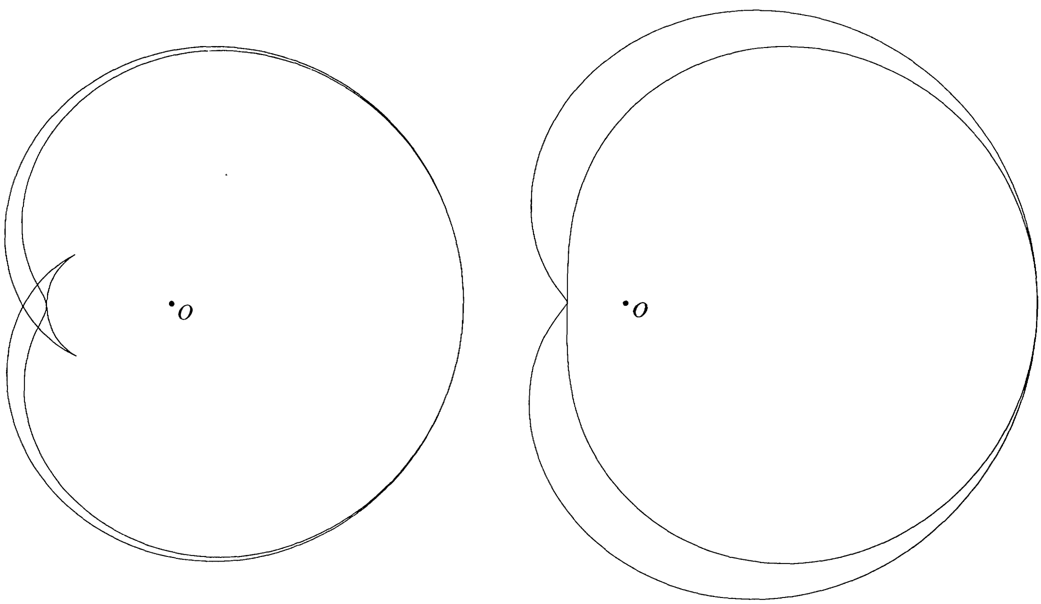


FIG. 16. Parametrizing oriented lines.

accurate picture of $CV(H)$ at such points. A more minor matter is that (r, w) and $(-r, -w)$ both project to the same point of $\mathbb{R}^2$: the projection identifies the two parts of $CV(H)$ to one curve in $\mathbb{R}^2$. It is a pleasant exercise to verify that the projection of $CV(H)$ is the *pedal curve* of C relative to $(0,0)$: the locus of the foot of the perpendicular from $(0,0)$ to the tangent to C. Here are two examples where the tangent to C never passes through $(0,0)$, so the pedal curve drawn is an accurate picture of the dual. See Fig. 17.

More about $CV(H)$ in § 6.



Limaçon with two inflexions, and its pedal curve. Limaçon with a higher inflexion, and its pedal curve.

FIG. 17.

(5.3) *Distance-squared functions*. Other families of functions are connected with other geometrical objects. For example let

$$D: S^1 \times \mathbb{R}^2 \to \mathbb{R}$$

be given by $D(t, w) = \|e(t) - w\|^2$: the distance-squared from $w \in \mathbb{R}^2$ to $e(t) \in \mathbb{R}^2$. We do not have space for all the details here (see [3] or [7]) but just mention the following. (i) For curves satisfying the generic property 4.9(ii) only A_1, A_2, A_3 singularities of $f(t) = D(t, w_0)$ occur, that is, we do not get $\partial^i D/\partial t^i(t_0, w_0) = 0$ for $i = 1,2,3,4$ except when $e(t_0)$ is a higher vertex of C

(and w_0 is the center of curvature of C at $e(t_0)$). (ii) $CV(D)$ is the surface in $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$ (coordinates (r, w)) whose section $r = \text{constant} \geqslant 0$ is the *parallels* to C at distance $\sqrt{r}$ from C. (iii) The "bifurcation set"

$$B(D) = \{w \in \mathbb{R}^2 \colon \text{for some } t,\ \partial D/\partial t = \partial^2 D/\partial t^2 = 0 \text{ at } (t, w)\}$$

is the *evolute* of C, the locus of centers of curvature. The proofs of these facts just use the standard formulae for plane curves connecting tangent, normal and curvature; see for example [9].

6. Versal Unfoldings. The family H of height functions on a curve satisfying the generic property 4.9(i) is itself generic in a sense to be explained below. The same goes for the family D of (5.3), assuming 4.9(ii). In the language of singularity theory H or D is called a *versal unfolding* and this has interesting geometrical consequences, which we consider first. To cover both H and D (not to mention other examples) we write our family as $F\colon S^1 \times W \to \mathbb{R}$ so that $W = S^1$ for $F = H$; $W = \mathbb{R}^2$ for $F = D$. Consider a specific point $(t_0, w_0) \in S^1 \times W$ and let $f(t) = F(t, w_0)$. As above, we shall say f has "type A_m at t_0" iff $f^{(i)}(t_0)$ is zero for $i = 1, \dots, m$ and nonzero for $i = m + 1$.

For F we define $CV(F)$ and $B(F)$ exactly as in the examples of § 5.

(6.1) *Suppose that F satisfies the genericity condition of being a "versal unfolding of f at t_0" (see below). Then the sets $CV(F)$ and $B(F)$ are determined locally, up to a smooth change of coordinates in $\mathbb{R} \times W$ and W respectively, by the number m.* □

By "locally" we mean that t, w and r are to be close to t_0, w_0 and $F(t_0, w_0)$ respectively. This amounts to saying that we consider for H (resp. D) only the part of the dual curve (resp. evolute) corresponding to points of C near $e(t_0)$.

The good thing about (6.1) is that we obtain an accurate picture of $CV(F)$ or $B(F)$, distorted only by smooth changes of coordinates, by considering any other versal unfolding with the same W and m. In practice we can, and do, choose very simple ones. It is clearly necessary to be able to recognize a versal unfolding when we see one; fortunately this is very easy, using the following standard result of singularity theory.

Write $w_1, \dots, w_n$ for local coordinates on W near w_0. Thus for height functions, $n = 1$ and w_1 = angle coordinate θ, so we regard S^1 near $w_0 = (\cos\theta_0, \sin\theta_0)$ as the same as $\mathbb{R}$ near θ_0. For distance-squared functions we use ordinary cartesian coordinates (w_1, w_2).

Given F, we can form $\partial F/\partial w_i(t, w_0)$ for $i = 1, \dots, n$. This is a function of t only, and we can write down its Taylor series at t_0, say

$$g_i = a_{i1}(t - t_0) + a_{i2}(t - t_0)^2 + \cdots + a_{i, m-1}(t - t_0)^{m-1}$$

where we omit the constant term and stop after $m - 1$ terms. There is the usual formula

$$a_{ij}(1/j!)\left(\partial^{j+1}F/\partial t^j \partial w_i\right)(t_0, w_0).$$

(6.2) *F is a versal unfolding* of f at t_0 iff the $n \times (m - 1)$ matrix of coefficients a_{ij} has rank $m - 1$.* □

Notes. (1) If $m = 1$, then the condition is vacuous and automatically satisfied.

(2) If $n \geqslant m - 1$, then we would expect "almost any" F to be a versal unfolding, for matrices have maximal rank unless the entries are unlucky enough to satisfy certain relations. Being a versal unfolding is a generic property, so long as $n \geqslant m - 1$.

(3) If $n < m - 1$, then F can never be versal.

(6.3) *Let $e\colon S^1 \to \mathbb{R}^2$ satisfy the generic property* (i) (*resp.* (ii)) *of* (4.9). *Then the family H of height functions* (*resp. D of distance-squared functions*) *is always a versal unfolding.*

Proof. We consider only the case of H and leave the reader to try the (slightly harder!) D. By

*Our family F is a family of "potential functions," that is, we ignore additive constants, as is appropriate for $F = H$ or $F = D$. There is a slightly different result when constants are important.

(5.2) we need only consider $m = 1$ and $m = 2$, and by note (1) we can stick to $m = 2$. Write

$$w = (\cos\theta, \sin\theta), \qquad e(t) = (X(t), Y(t))$$

so that

$$H(t,\theta) = X(t)\cos\theta + Y(t)\sin\theta.$$

Thus

$$\partial H/\partial\theta = -X(t)\sin\theta + Y(t)\cos\theta$$

and

$$g_1 = (-X'(t_0)\sin\theta_0 + Y'(t_0)\cos\theta_0)(t - t_0).$$

We just need to know that the coefficient of $(t - t_0)$ is nonzero. Now from $m = 2$ we know $\partial H/\partial t(t_0, \theta_0) = 0$ (also $\partial^2 H/\partial t^2 = 0$, but we don't need to use that), so

$$X'(t_0)\cos\theta_0 + Y'(t_0)\sin\theta_0 = 0.$$

Since $X'(t_0)$ and $Y'(t_0)$ won't both be zero ($e'(t)$ is never 0 for an embedding e) this easily implies that the required coefficient is nonzero. □

Consider the family $F(t,\theta) = t^2$, for which $n = 1$, and for this family take $t_0 = 0$, $\theta_0 = 0$ (so $w_0 = (1,0)$). Then $f(t) = F(t,0)$ has type A_1 ($m = 1$) at 0 and this is automatically a versal unfolding. Further

$$CV(F) = \{(r,\theta)\colon \text{for some } t,\ t = 0 \text{ and } t^2 = r\}$$

and close to $(0,0)$ this is just a little interval of the θ-axis: the local picture of $CV(F)$ is a piece of straight line in a plane. A smooth change of coordinates in the plane will take this to a piece of smooth curve (Fig. 18).

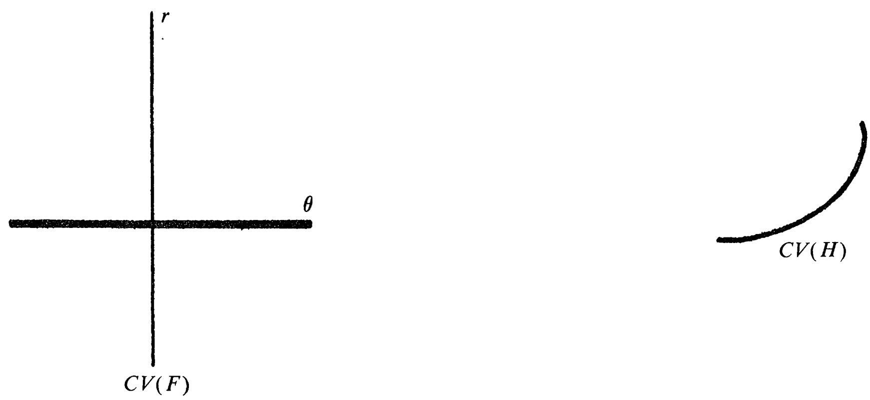


FIG. 18. Critical value set, type A_1—smooth.

For the family H of height functions on a curve C, the condition $m = 1$ means that C does *not* have an inflexion at the relevant point $e(t_0)$ (see (5.2)). Thus the part of C close to $e(t_0)$ has dual which is a piece of smooth curve, by (6.1).

More interesting is the case $m = 2$. For this consider $F(t,\theta) = t^3 + \theta t$, where again we take $t_0 = \theta_0 = 0$ for this F. Certainly $f(t) = F(t,0) = t^3$ has type A_2 ($m = 2$) at 0; also $\partial F/\partial\theta = t$, so the matrix of (6.2) is the 1×1 matrix (1). Hence F is versal. Further

$$CV(F) = \{(r,\theta)\colon 3t^2 + \theta = 0,\ t^3 + \theta t = r \text{ for some } t\}$$

which comes to $\{(r,\theta)\colon 27r^2 + 4\theta^3 = 0\}$. Any smooth change of coordinates takes this to what is called an *ordinary cusp* in the plane (Fig. 19). Since, for height functions on C, $m = 2$ means that the curve C has an ordinary inflexion at the relevant point $e(t_0)$ (see (5.2)) we have:

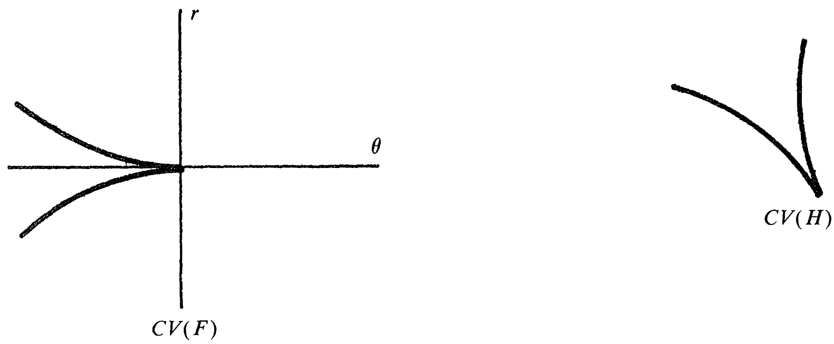


FIG. 19. Critical value set, type A_2—cusp.

(6.4) *For a plane curve C satisfying the generic property* 4.9(i) *the dual (or the pedal away from* (0, 0)) *is a smooth curve except for the presence of ordinary cusps corresponding with ordinary inflexions of C.* □

Compare Fig. 17, noting that there may be self-crossings on the dual, corresponding to double tangents to C. These are not part of the *local* structure: each branch at the crossing is smooth.

A similar analysis using the distance-squared functions D and the bifurcation set $B(D)$ shows the following.

(6.5) *For a plane curve C satisfying the generic property* 4.9(ii) *the evolute is a smooth curve except for the presence of ordinary cusps corresponding with ordinary vertices of C.* □

Adaptations of the methods above show that the property 4.9(i) together with the absence of tritangents is generic (but not the latter by itself—can you see why?). This implies that duals generically have no *triple* self-crossings. Double self-crossings, as in Fig. 17, are stable under small perturbations of C, but triple crossings split up into three double crossings.

Other families of functions are connected with the study of *envelopes*, and one can study singular points (e.g., cusps) of these envelopes by analogous methods. We give a good deal of detail on this in [7].

References

1. V. I. Arnold, Wavefront evolution and equivariant Morse lemma, Comm. Pure Appl. Math., 29 (1976) 557–582.

2. ______, Indices of singular points of 1-forms on a manifold with boundary, convolution of invariants of reflection groups, and singular projections of smooth surfaces, Russian Math Surveys, 34 (1979) 1–42.

3. T. Banchoff, T. Gaffney, and C. McCrory, Cusps of Gauss Mappings, Research Notes in Mathematics, 55, Pitman, 1982.

4. Th. Bröcker and L. Lander, Differentiable Germs and Catastrophes, London Math. Soc. Lecture Notes, 17, Cambridge Univ. Press, 1975.

5. J. W. Bruce, The duals of generic hypersurfaces, Math. Scand., 49 (1981) 36–60.

6. ______, Wavefronts and parallels in Euclidean space, to appear in Math. Proc. Cambridge Philos. Soc.

7. J. W. Bruce and P. J. Giblin, Curves and Singularities, in press, Cambridge Univ. Press.

8. J. W. Bruce, P. J. Giblin, and C. G. Gibson, On caustics by reflexion, Topology, 21 (1982) 179–199.

9. M. P. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.

10. V. Guillemin and A. Pollack, Differential Topology, Prentice-Hall, 1974.

11. E. J. N. Looijenga, Structural stability of smooth families of C^∞ functions, Thesis, University of Amsterdam, 1974.

12. J. W. Milnor, Topology from the Differentiable Viewpoint, Univ. of Virginia Press, 1965.

13. I. R. Porteous, The normal singularities of a submanifold, J. Differential Geom., 5 (1971) 543–564.

14. C. T. C. Wall, Geometric Properties of Generic Differentiable Manifolds, Springer Lecture Notes in Math., 597 (1977) 707–774.

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

L. F. KLOSINSKI
Department of Mathematics, University of Santa Clara, CA 95053

G. L. ALEXANDERSON
Department of Mathematics, University of Santa Clara, CA 95053

A. P. HILLMAN
Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131

The following results of the forty-third William Lowell Putnam Mathematical Competition, held on December 4, 1982, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, five thousand dollars, was awarded to the Department of Mathematics of **Harvard University**, Cambridge, Massachusetts. The members of its winning team were: Benji N. Fisher, Michael J. Larsen, and Michael Raship; each was awarded a prize of two hundred fifty dollars.

The second prize, two thousand five hundred dollars, was awarded to the Department of Mathematics of the **University of Waterloo**, Waterloo, Ontario, Canada. The members of its team were: David W. Ash, W. Ross Brown, and Herbert J. Fichtner; each was awarded a prize of two hundred dollars.

The third prize, one thousand five hundred dollars, was awarded to the Department of Mathematics of the **California Institute of Technology**, Pasadena, California. The members of its team were: Bradley W. Brock, Scott R. Johnson, and Zinovy B. Reichstein; each was awarded a prize of one hundred fifty dollars.

The fourth prize, one thousand dollars, was awarded to the Department of Mathematics of **Yale University**, New Haven, Connecticut. The members of its team were Alan S. Edelman, Paul N. Feldman, and Nathaniel E. Glasser; each was awarded a prize of one hundred dollars.

The fifth prize, five hundred dollars, was awarded to the Department of Mathematics of **Princeton University**, Princeton, New Jersey. The members of its team were Gregg N. Patruno, David P. Roberts, and Daniel J. Scales; each was awarded a prize of fifty dollars.

The five highest-ranking individual contestants, in alphabetical order, were **David W. Ash**, University of Waterloo; **Eric D. Carlson**, Michigan State University; **Noam D. Elkies**, Columbia University; **Brian R. Hunt**, University of Maryland, College Park; and **Edward A. Shpiz**, Washington University, St. Louis. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of five hundred dollars by the Putnam Prize Fund.

The next six highest-ranking individuals, in alphabetical order, were *Joel Friedman*, Harvard University; *Irwin L. Jungreis*, Cornell University; *Michael J. Larsen*, Harvard University; *Gregg N. Patruno*, Princeton University; *Robin A. Pemantle*, University of California, Berkeley; and *Mark G. Pleszkoch*, University of Virginia. Each of these students was awarded a prize of two hundred fifty dollars.

The following teams, named in alphabetical order, received honorable mention: *Brown University*, with team members Stephen A. DiPippo, Edward F. Grove, and Erik R. Paulsen; *University of California, Berkeley*, with team members John W. Jones, Robin A. Pemantle, and Richard W. Webb; *Michigan State University*, with team members Eric D. Carlson, Erin J. Schram, and Frank J. Sottile; *University of Toronto*, with team members John J. Chew, John J. Im, Alastair M. Rucklidge; and *Washington University*, St. Louis, with team members Bard Bloom, Edward A. Shpiz, and Richard A. Stong.

Honorable mention was achieved by the following thirty-three individuals, named in alphabetical order: *Gary M. Bernstein*, Princeton University; *Bradley W. Brock*, California Institute of Technology; *W. Ross Brown*, University of Waterloo; *Bev I. Cope*, University of Waterloo; *Charles J. Cuny*, California Institute of Technology; *Michael J. Dawson*, Massachusetts Institute of Technology; *David L. Desjardins*, Massachusetts Institute of Technology; *Stephen A. DiPippo*, Brown University; *Paul N. Feldman*, Yale University; *Michael V. Finn*, Harvard University; *Zachary M. Franco*, Harvard University; *Tom Ilmanen*, Haverford College; *John J. Im*, University of Toronto; *Scott R. Johnson*, California Institute of Technology; *Kin Y. Li*, University of Washington; *F. Miller Maley*, Amherst College; *Yair N. Minsky*, Princeton University; *Evan W. Morton*, Harvard University; *Alan G. Murray*, California Institute of Technology; *Laurence E. Penn*, Harvard University; *Jeremy D. Primer*, Princeton University; *Michael Raship*, Harvard University; *David P. Roberts*, Princeton University; *James R. Roche*, University of Notre Dame; *Richard A. Shapiro*, Massachusetts Institute of Technology; *Carlos T. Simpson*, Harvard University; *Arthur P. Smith*, Memorial University of Newfoundland; *Thomas R. Stevenson*, University of British Columbia; *Richard A. Stong*, Washington University, St. Louis; *John M. Sullivan*, Harvard University; *Jerome V. Walsh*, University of Illinois at Urbana-Champaign; *Dwight S. Wilson*, Johns Hopkins University; and *David Wolland*, Harvard University.

The other individuals who achieved ranks among the top 100, in alphabetical order of their schools, were: University of Alberta, *Arthur B. Baragar*, *Robert P. Morewood*; University of Arizona, *Jan M. O. Soderkvist*; Biola University, *Mark M. Shimozono*; University of British Columbia, *Edmond D. Chow*; Brown University, *Erik R. Paulsen*; California Institute of Technology, *Pang-Chieh Chen*, *R. Sekhar Chivukula*, *James T. Lin*, *Vladimir S. Matijasevic*, *Zinovy B. Reichstein*; University of California at Santa Barbara, *John R. Kelly*; Carleton University, *Mike D. Dixon*; Case Western Reserve University, *Kevin E. Kelso*; University of Chicago, *Keith A. Ramsay*, *Yongbum Park*, *Michael P. Spertus*, *David S. Yuen*; University of Colorado, Boulder, *Boris Lerner*; Harvard University, *Frederick R. Adler*, *Bruce W. K. Brandt*, *Benji N. Fisher*, *Howard M. Pollack*, *Gregory B. Sorkin*; Haverford College, *Samuel R. Evens*; Iowa State University, *William R. Somsky*; Université Laval, *Line Baribeau*; University of Maryland, College Park, *Dougin A. Walker*; Massachusetts Institute of Technology, *David E. Brahm*, *Andrew E. Gelman*, *Daniel S. Lewart*, *Tomasz S. Mrowka*; Michigan State University, *Frank J. Sottile*; University of Michigan, Ann Arbor, *David A. Short*; University of Michigan, Dearborn, *Gregory T. Parker*; University of New Brunswick, *Christian Friesen*; North Carolina State University, *Samuel P. White*; Pomona College, *Alan M. Nadel*; Princeton University, *Kazuko Suzuki*, *Burt J. Totaro*, *Kevin M. Walker*; Queen's College of the City University of New York, *Boris Aronov*; Reed College, *Paul S. Hsieh*; Rensselaer Polytechnic Institute, *Benjamin J. Patz*, *William J. Harte*; Rose-Hulman Institute of Technology, *Randy L. Ekl*; Rutgers University, *Scott E. Axelrod*; Simon Fraser University, *David G. Wagner*; Sioux Falls College, *William H. Paulsen*; Stanford University, *Cris G. Poor*; University of Toronto, *John J. Chew*, *Peter A. Miegom*; Washington University, St. Louis, *Patrick J. Abegg*, *Paul H. Burchard*; University of Waterloo, *Herbert S. Fichtner*, *Charles R. Smith*, *Charles S. A. Timar*; University of Wisconsin, *Jon G. Udell*; Yale University, *Alan S. Edelman*, *Nathaniel E. Glasser*.

There were 2024 individual contestants from 348 colleges and universities in Canada and the United States in the competition of December 4, 1982. Teams were entered by 249 institutions.

The Questions Committee for the forty-third competition consisted of William J. Firey (Chairman), Douglas A. Hensley, and Melvin Hochster; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1

Let V be the region in the cartesian plane consisting of all points (x, y) satisfying the simultaneous conditions

$$|x| \leqslant y \leqslant |x| + 3 \quad \text{and} \quad y \leqslant 4.$$

Find the centroid $(\bar{x}, \bar{y})$ of V.

Problem A-2

For positive real x, let

$$B_n(x) = 1^x + 2^x + 3^x + \cdots + n^x.$$

Prove or disprove the convergence of

$$\sum_{n=2}^{\infty} \frac{B_n(\log_n 2)}{(n \log_2 n)^2}.$$

Problem A-3

Evaluate

$$\int_0^{\infty} \frac{\operatorname{Arctan}(\pi x) - \operatorname{Arctan} x}{x}\, dx.$$

Problem A-4

Assume that the system of simultaneous differential equations

$$y' = -z^3, \quad z' = y^3$$

with the initial conditions $y(0) = 1$, $z(0) = 0$ has a unique solution $y = f(x)$, $z = g(x)$ defined for all real x. Prove that there exists a positive constant L such that for all real x,

$$f(x + L) = f(x), \quad g(x + L) = g(x).$$

Problem A-5

Let a, b, c, and d be positive integers and

$$r = 1 - \frac{a}{b} - \frac{c}{d}.$$

Given that $a + c \leqslant 1982$ and $r > 0$, prove that

$$r > \frac{1}{1983^3}.$$

Problem A-6

Let σ be a bijection of the positive integers, that is, a one-to-one function from $\{1, 2, 3, \dots\}$ onto itself. Let $x_1, x_2, x_3, \dots$ be a sequence of real numbers with the following three properties:

(i) $|x_n|$ is a strictly decreasing function of n;
(ii) $|\sigma(n) - n| \cdot |x_n| \to 0$ as $n \to \infty$;
(iii) $\lim_{n\to\infty} \sum_{k=1}^{n} x_k = 1$.

Prove or disprove that these conditions imply that

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_{\sigma(k)} = 1.$$

Problem B-1

Let M be the midpoint of side BC of a general ΔABC. Using the *smallest possible* n, describe a method for cutting ΔAMB into n triangles which can be reassembled to form a triangle congruent to ΔAMC.

Problem B-2

Let $A(x, y)$ denote the number of points (m, n) in the plane with integer coordinates m and n satisfying

$m^2 + n^2 \leqslant x^2 + y^2$. Let $g = \Sigma_{k=0}^{\infty} e^{-k^2}$. Express

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x, y) e^{-x^2-y^2}\, dx\, dy$$

as a polynomial in g.

Problem B-3

Let p_n be the probability that $c + d$ is a perfect square when the integers c and d are selected independently at random from the set $\{1, 2, 3, \ldots, n\}$. Show that $\lim_{n\to\infty}(p_n\sqrt{n})$ exists and express this limit in the form $r(\sqrt{s} - t)$, where s and t are integers and r is a rational number.

Problem B-4

Let $n_1, n_2, \ldots, n_s$ be distinct integers such that

$$(n_1 + k)(n_2 + k) \cdots (n_s + k)$$

is an integral multiple of $n_1 n_2 \cdots n_s$ for every integer k. For each of the following assertions, give a proof or a counterexample:

(a) $|n_i| = 1$ for some i.
(b) If further all n_i are positive, then

$$\{n_1, n_2, \ldots, n_s\} = \{1, 2, \ldots, s\}.$$

Problem B-5

For each $x > e^e$ define a sequence $S_x = u_0, u_1, u_2, \ldots$ recursively as follows: $u_0 = e$, while for $n \geqslant 0$, u_{n+1} is the logarithm of x to the base u_n. Prove that S_x converges to a number $g(x)$ and that the function g defined in this way is continuous for $x > e^e$.

Problem B-6

Let $K(x, y, z)$ denote the area of a triangle whose sides have lengths x, y, and z. For any two triangles with sides a, b, c and a', b', c', respectively, prove that

$$\sqrt{K(a, b, c)} + \sqrt{K(a', b', c')} \leqslant \sqrt{K(a + a', b + b', c + c')}$$

and determine the cases of equality.

SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geqslant i \geqslant 0$ is the number of students among the top 201 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1. $(146, 9, 2, 5, 0, 0, 0, 5, 2, 9, 22, 1)$

Let T consist of the points inside or on the triangle with vertices at $(0,3), (-1,4), (1,4)$ and let U be the set of points inside or on the triangle with vertices at $(0,0), (-4,4), (4,4)$. Then T and V overlap only on boundary points and their union is U. The centroids of T and U are $(0, 11/3)$ and $(0, 8/3)$, respectively. The areas of T, V, and U are 1, 15, and 16, respectively. Using weighted averages with the areas as weights, one has

$$1 \cdot 0 + 15\bar{x} = 16 \cdot 0, \qquad 1 \cdot \tfrac{11}{3} + 15\bar{y} = 16 \cdot \tfrac{8}{3}.$$

It follows that $\bar{x} = 0$, $\bar{y} = 13/5$.

A-2. $(68, 17, 18, 7, 0, 3, 1, 0, 10, 2, 37, 38)$

Since $x = \log_n 2 > 0$, $B_n(x) = 1^x + 2^x + \cdots + n^x \leqslant n \cdot n^x$ and

$$0 \leqslant \frac{B_n(\log_n 2)}{(n\log_2 n)^2} \leqslant \frac{n \cdot n^{\log_n 2}}{(n\log_2 n)^2} = \frac{2}{n(\log_2 n)^2}.$$

As $\Sigma_{n=2}^{\infty}[2/n(\log_2 n)^2]$ converges by the integral test, the given series converges by the comparison test.

A-3. (14, 19, 11, 5, 0, 0, 0, 2, 5, 0, 37, 108)

$$\begin{aligned}\int_0^{\infty}\frac{\operatorname{Arctan}(\pi x)-\operatorname{Arctan}x}{x}\,dx &= \int_0^{\infty}\frac{1}{x}\operatorname{Arctan}(ux)\Big|_{u=1}^{u=\pi}dx\\ &= \int_0^{\infty}\int_1^{\pi}\frac{1}{1+(xu)^2}\,du\,dx=\int_1^{\pi}\int_0^{\infty}\frac{1}{1+(xu)^2}\,dx\,du\\ &= \int_1^{\pi}\frac{1}{u}\cdot\frac{\pi}{2}\,du=\frac{\pi}{2}\ln\pi.\end{aligned}$$

A-4. (0, 2, 4, 5, 0, 0, 0, 11, 9, 22, 33, 115)

The differential equations imply that

$$y^3y' + z^3z' = z'y' - y'z' = 0$$

and hence that $y^4 + z^4$ is constant. This and the initial conditions give $y^4 + z^4 = 1$. Thinking of x as a time variable and (y, z) as the coordinates of a point in a plane, this point moves on the curve $y^4 + z^4 = 1$ with speed

$$\left[(y')^2+(z')^2\right]^{1/2}=\left(z^6+y^6\right)^{1/2}.$$

At any time, either y^4 or z^4 is at least $\frac{1}{2}$ and so the speed is at least $\{[(\frac{1}{2})^{1/4}]^6\}^{1/2}$. Hence the point will go completely around the finite curve in some time L. As the speed depends only on y and z (and not on x), the motion is periodic with period L.

A-5. (22, 7, 0, 1, 0, 0, 0, 0, 5, 4, 48, 114)

We are given that

$$r = 1-\frac{a}{b}-\frac{c}{d}=\frac{bd-ad-bc}{bd}>0.$$

Thus $bd - ad - bc$ is a positive integer and so $r \geqslant 1/bd$. We may assume without loss of generality that $b \leqslant d$. If $b \leqslant d \leqslant 1983$, $r \geqslant 1983^{-2} > 1983^{-3}$. Since $a + c \leqslant 1982$, if $1983 \leqslant b \leqslant d$, one has

$$r\geqslant 1-\frac{a}{1983}-\frac{c}{1983}\geqslant 1-\frac{1982}{1983}=\frac{1}{1983}>\frac{1}{1983^3}.$$

The remaining case is that with $b < 1983 < d$. Then the d that minimizes r for fixed a, b, c is $1 + [bc/(b - a)]$, where $[x]$ is the greatest integer in x. This d is at most $1983b$ since $b - a \geqslant 1$ and $c < 1982$ and thus

$$r\geqslant\frac{1}{bd}\geqslant\frac{1}{1983b^2}>\frac{1}{1983^3}.$$

Hence we have the desired inequality in all cases.

A-6. (0, 0, 0, 0, 4, 1, 2, 0, 5, 4, 50, 135)

We disprove the assertion. Let $y_n = 1/(n + 1)\ln(n + 1)$. Then $\Sigma_{n=1}^{\infty}(-1)^{n+1}y_n$ converges to some $g > 0$ since $y_n \to 0$ as $n \to \infty$ and $y_1 > y_2 > \cdots$. Let $x_n = (-1)^{n+1}y_n/g$. Then conditions (i) and (iii) are satisfied. Let $a_0, a_1, \ldots$ be positive integers to be made more definite later. Let $b_0 = 0$ and $b_{i+1} = b_i + 4a_i$ for $i = 0, 1, \ldots$. The bijection σ is defined as follows:

$$\sigma(n) = 2n - 1 - b_i \quad \text{for} \quad b_i < n \leqslant b_i + 2a_i,$$

$$\sigma(n) = 2n - b_{i+1} \quad \text{for} \quad b_i + 2a_i < n \leqslant b_{i+1}.$$

Then $0 < \sigma(n) < 2n$ and hence $|\sigma(n) - n| < n$. Thus

$$|\sigma(n) - n| \cdot |x_n| < \frac{n}{g(n+1)\ln(n+1)}$$

which implies condition (ii). Let $C(n) = \sum_{i=1}^{n} x_i$ and $D(n) = \sum_{i=1}^{n} x_{\sigma(i)}$. Then

$$\text{(A)} \qquad D(b_i + 2a_i) - C(b_i + 2a_i) = \frac{1}{g}\sum_{j=1}^{a_i} y_{b_i+2j} + \frac{1}{g}\sum_{k=1}^{a_i} y_{b_i+2a_i+2k-1}.$$

Since $y_2 + y_4 + y_6 + \cdots$ diverges to $+\infty$ by the integral test, the a_i can be chosen large enough so that the first sum in (A) exceeds 1 for each i. Then $D(n) > 1 + C(n)$ for an unbounded sequence of n's. Hence $D(n)$ and $C(n)$ cannot converge to the same limit.

B-1. (56, 46, 23, 0, 0, 0, 0, 0, 18, 5, 5, 48)

The smallest n is 2. Let D be the midpoint of side AB. Cut ΔAMB along DM. Then ΔBMD can be placed alongside ΔADM, with side BD atop side AD, so as to form a triangle congruent to ΔAMC. Since ΔAMB need not be congruent to ΔAMC in a general ΔABC, there is no method with $n = 1$.

B-2. (5, 28, 16, 3, 2, 0, 0, 0, 4, 8, 32, 103)

Let $r = \sqrt{x^2 + y^2}$, $R(m, n) = \{(x, y) : m^2 + n^2 \leqslant x^2 + y^2\}$, and

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x, y)e^{-x^2-y^2}\,dx\,dy.$$

Let Σ and Σ' denote sums over all integers m and over all integers n, respectively. Then

$$\begin{aligned}
I &= \sum\sum{}'\iint_{R(m,n)} e^{-x^2-y^2}\,dx\,dy \\
&= \sum\sum{}'\int_0^{2\pi}\int_{\sqrt{m^2+n^2}}^{\infty} e^{-r^2} r\,dr\,d\theta \\
&= \sum\sum{}'\int_0^{2\pi}\left[-\tfrac{1}{2}e^{-r^2}\right]_{\sqrt{m^2+n^2}}^{\infty} d\theta \\
&= \sum\sum{}'\int_0^{2\pi}\tfrac{1}{2}e^{-m^2-n^2}\,d\theta \\
&= \sum\sum{}'\pi e^{-m^2-n^2} = \pi\left(\sum e^{-m^2}\right)\left(\sum{}' e^{-n^2}\right) = \pi(2g-1)^2.
\end{aligned}$$

B-3. (82, 18, 5, 0, 0, 0, 0, 3, 8, 6, 40, 39)

Let $a(n) = [\sqrt{n+1}\,]$ and $b(n) = [\sqrt{2n}\,]$, where $[x]$ is the greatest integer in x. For t in $\{1, 2, \ldots, a(n)\}$, there are $t^2 - 1$ ordered pairs (c, d) with c and d in $X = \{1, 2, \ldots, n\}$ and $c + d = t^2$. For t in $\{1 + a(n), 2 + a(n), \ldots, b(n)\}$, there are $2n + 1 - t^2$ ordered pairs (c, d) with c and d in X and $c + d = t^2$. Hence the total number $F(n)$ of favorable (c, d) is

$$\begin{aligned}
F(n) &= \sum_{t=1}^{a(n)} (t^2 - 1) + \sum_{t=1+a(n)}^{b(n)} (2n + 1 - t^2) \\
&= \left(2\sum_{t=1}^{a(n)} t^2\right) - \left(\sum_{t=1}^{b(n)} t^2\right) - a(n) + [b(n) - a(n)](2n+1) \\
&= \frac{2a(n)[1 + a(n)][1 + 2a(n)]}{6} - \frac{b(n)[1 + b(n)][1 + 2b(n)]}{6} \\
&\quad - 2(n+1)a(n) + (2n+1)b(n).
\end{aligned}$$

Since $p_n = F(n)/n^2$,

$$\lim_{n\to\infty}\left(p_n\sqrt{n}\right) = \lim_{n\to\infty} F(n)/n^{3/2}$$

$$= \frac{2\cdot 2}{6}\lim_{n\to\infty}\left(\frac{a(n)}{\sqrt{n}}\right)^3 - \frac{2}{6}\lim_{n\to\infty}\left(\frac{b(n)}{\sqrt{n}}\right)^3 - 2\lim_{n\to\infty}\frac{a(n)}{\sqrt{n}} + 2\lim_{n\to\infty}\frac{b(n)}{\sqrt{n}}$$

$$= \frac{2}{3} - \frac{1}{3}(\sqrt{2})^3 - 2 + 2\sqrt{2} = \frac{4}{3}(\sqrt{2}-1).$$

B-4. (2, 7, 5, 1, 13, 3, 3, 1, 0, 2, 36, 128)

Let $P_k = (n_1 + k)(n_2 + k)\cdots(n_s + k)$. We are given that $P_0|P_k$ for all integers k.

(a) $P_0|P_{-1}$ and $P_0|P_1$ together imply $P_0^2|(P_{-1}P_1)$ or $(n_1^2n_2^2\cdots n_s^2)|[(n_1^2-1)(n_2^2-1)\cdots(n_s^2-1)]$.

No n_i can be zero since $P_k \neq 0$ for k sufficiently large. Thus, for each i, $n_i^2 \geqslant 1$ and $n_i^2 > n_i^2 - 1 \geqslant 0$. Hence $P_0^2 > P_{-1}P_1 \geqslant 0$. This and $P_0^2|(P_{-1}P_1)$ imply $P_{-1}P_1 = 0$. Then for some i, $|n_i| = 1$.

(b) P_k is a polynomial in k of degree s. Since P_0 divides each P_i, P_0 also divides the nth difference

$$\sum_{i=0}^{s}(-1)^i\binom{s}{i}P_i = s!.$$

Since $P_0 > 0$, this means that $P_0 \leqslant s!$. As P_0 is a product of s distinct positive integers, it follows that

$$\{n_1, n_2, \ldots, n_s\} = \{1, 2, \ldots, s\}.$$

B-5. (2, 3, 1, 2, 3, 1, 0, 4, 9, 23, 44, 109)

Since the derivative of $x^{1/x}$ is negative for $x > e$,

(1) $$a^b > b^a \text{ when } e \leqslant a < b.$$

The u's are defined so that $u_0 = e$ and

(2) $$x = (u_0)^{u_1} = (u_1)^{u_2} = (u_2)^{u_3} = \cdots.$$

Hence

(3) $$u_{n+1} = (u_n \ln u_{n-1})/\ln u_n.$$

As $x > e^e$, $u_1 = \ln x > e = u_0$. Now $u_1 > u_0$ implies $\ln u_1 > \ln u_0$ and then (3) with $n = 1$ implies $u_2 < u_1$. Also, (2) and (1) imply $(u_1)^{u_2} = (u_0)^{u_1} > (u_1)^{u_0}$, which gives us $u_2 > u_0$. Now $u_2 < u_1$ and (3) with $n = 2$ imply $u_3 > u_2$. Also (2) and (1) imply $(u_2)^{u_3} = (u_1)^{u_2} < (u_2)^{u_1}$ and hence $u_3 < u_1$. Similarly, $u_2 < u_4 < u_3$. Then an easy induction shows that

$$e < u_{2n} < u_{2n+2} < u_{2n+1} < u_{2n-1} \text{ for } n = 1, 2, \ldots.$$

Thus the monotonic bounded sequences $u_0, u_2, u_4, \ldots$ and $u_1, u_3, u_5, \ldots$ have limits a and b, respectively, with $e < a \leqslant b$. Also

$$a^b = \lim_{n\to\infty}(u_{2n})^{u_{2n+1}} = x = \lim_{n\to\infty}(u_{2n-1})^{u_{2n}} = b^a.$$

Then $a^b = b^a$, $e \leqslant a \leqslant b$, and (1) imply $a = b$. Hence $\lim_{n\to\infty} u_n$ exists and is the unique real number $g = g(x)$ with $g > e$ and $g^g = x$. Since $f(y) = y^y$ is continuous and strictly increasing for $y \geqslant e$, its inverse function $g(x)$ is also continuous.

B-6. (10, 3, 7, 0, 0, 0, 0, 0, 1, 8, 10, 29, 133)

Let $s = (a + b + c)/2$, $t = s - a$, $u = s - b$, $v = s - c$ and similarly for the primed letters.

Using Heron's Formula, the inequality to be proved will follow from

$$\text{(A)} \qquad \sqrt[4]{stuv} + \sqrt[4]{s't'u'v'} \leqslant \sqrt[4]{(s+s')(t+t')(u+u')(v+v')}$$

for positive $s, t, u, v, s', t', u', v'$. A simpler analogous inequality that might be helpful is

$$\text{(B)} \qquad \sqrt{xy} + \sqrt{x'y'} \leqslant \sqrt{(x+x')(y+y')} \quad \text{for} \quad x, y, x', y' \text{ positive.}$$

First we note that (B) follows from the Cauchy Inequality applied to the vectors $(\sqrt{x}, \sqrt{x'})$ and $(\sqrt{y}, \sqrt{y'})$ [and also follows from $(\sqrt{xy'} - \sqrt{x'y})^2 \geqslant 0$ or from the Inequality on the Means applied to xy' and $x'y$]. Using (B) with $x = \sqrt{st}$, $x' = \sqrt{s't'}$, $y = \sqrt{uv}$, $y' = \sqrt{u'v'}$ and reapplying (B) to the new right side, one has

$$\begin{aligned} \sqrt[4]{stuv} + \sqrt[4]{s't'u'v'} &\leqslant \sqrt{(\sqrt{st} + \sqrt{s't'})(\sqrt{uv} + \sqrt{u'v'})} \\ &\leqslant \sqrt{\sqrt{(s+s')(t+t')}\sqrt{(u+u')(v+v')}}\,. \end{aligned}$$

Since here the rightmost part equals the right side of (A), we have proved (A).

Equality holds in (B) if and only if $\sqrt{x} : \sqrt{x'} = \sqrt{y} : \sqrt{y'}$ and this holds if and only if $x : x' = y : y'$. Hence equality occurs in (A) if and only if $s : t : u : v = s' : t' : u' : v'$. It follows that equality occurs in the original inequality if and only if a, b, c are proportional to a', b', c'.

PROGRESS REPORTS

EDITED BY THOMAS BANCHOFF AND RICHARD MILLMAN

It is easy to be too busy to pay attention to what anyone else is doing, but not good. All of us should know, and want to know, what has been discovered since our formal education ended, but new words, and relations between them, are growing too fast to keep up. It is possible for a person to learn of the title of a recent work and of the key words used in it and still not have the faintest idea of what the subject is.

Progress Reports is to be an almost periodic column intended to increase everyone's mathematical information about what others have been up to. Each column will report one step forward in the mathematics of our time. The purpose is to inform, more than to instruct: what is the name of the subject, what are some of the words it uses, what is a typical question, what is the answer, who found it. The emphasis will be on concrete questions and answers (theorems), and not on general contexts and techniques (theories). References will be kept minimal: usually they will include only one of the earliest papers in which the answer appears and a more recent exposition of the discovery, whenever one is easily available.

Everyone is invited to nominate subjects to be reported on and authors to prepare the reports. The ground rules are that the principal theorem should be old enough to have been published in the usual sense of that word (and not just circulated by word of mouth or in preprints); it should be of interest to more than just a few specialists; and it should be new enough to have an effect on the mathematical life of the present and near future. In practice most reports will probably be on progress achieved somewhere between 5 and 15 years ago.

MANIFOLDS WITH THE SAME SPECTRUM

RICHARD S. MILLMAN

Department of Mathematical & Computer Sciences, Michigan Technological University, Houghton, MI 49931

Many extremely important results in differential geometry concern the effect of the geometry of an object upon its topology. In addition, over the past forty years there has been a great deal of interest in the relationship between analytic quantities and the geometry and topology of

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P, L. A Handbook of Game Design. Henry Ellington, Eric Addinall, Fred Percival. Nichols Pub, 1982, 156 pp, $25. [ISBN: 0-89397-134-0] A systematic design strategy for developing card, board, computer and other types of games. Most of the examples discussed either are educational or mimic real life situations (war, business, politics, etc.), with no discussion of mathematical game theory, winning strategies, etc. Bibliography, glossary, extended case study, tips on holding competitions. GHM

Precalculus, T(13: 1). Algebra and Trigonometry, Third Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1983, 632 pp, $23.95. [ISBN: 0-673-15794-6] Standard topics in precalculus. Changes from Second Edition include more thorough review of algebra, and new sections on complex numbers and polar coordinates. (TR, First Edition, October 1978; Second Edition, December 1980.) JRG

Precalculus, T(13: 1). Essentials of Trigonometry. Karl J. Smith. Brooks/Cole Pub, 1983, xi + 292 pp, $19.95. [ISBN: 0-534-01224-8] In response to the "back-to-basics" movement the development here begins with right triangles but recognizes modern technology with extensive use of hand-held calculators. Standard topics, routine treatment; includes complex numbers, polar coordinates, logarithms, tables. JS

Education, S(15-16). Courseware in the Classroom: Selecting, Organizing, and Using Educational Software. Ann Lathrop, Bobby Goodson. Addison-Wesley, 1983, viii + 187 pp, $10 (P). [ISBN: 0-20007-4] Practical tips for the neophyte. Includes ideas on evaluating, sources of software and software reviews, and a courseware directory (to be updated annually). Illustrated with photos of actual programs on-screen. Examples are primarily elementary level in all subject areas. Good supplement for a course on using computers to teach mathematics. MW

Education, P, L*. The Agenda in Action: 1983 Yearbook. Gwen Shufelt, James R. Smart. NCTM, 1983, x + 245 pp, $14.50. [ISBN: 0-87353-201-5] 27 papers dealing with specific strategies for implementing eight of the recommendations in NCTM's 1980 pamphlet Agenda for Action (TR, December 1980). Topics range from problem solving to statistics, from computers to basic skills. A rich source of the best current thinking on mathematics education. LAS

History, P. Mathematical Perspectives: Essays on Mathematics and Its Historical Development. Ed: Joseph W. Dauben. Academic Pr, 1981, xv + 272 pp, $34. [ISBN: 0-12-204050-3] 13 scholarly essays in German (7), English (5) and French (1) on the history of mathematics, mostly on topics or people of the 17th, 18th and 19th centuries: Gauss, Kummer, Crelle, analytic functions, isoperimetric figures. A supplement to Historia Mathematica dedicated to Kurt-Reinhard Biermann on this 60th birthday. LAS

History, P, L*. P.R. Halmos: Selecta Expository Writing. Ed: Donald E. Sarason, Leonard Gillman. Springer-Verlag, 1983, xix + 304 pp, $19.80. [ISBN: 0-387-90756-4] Second volume of Halmos' Selecta, this one contains 27 expository gems ranging from Monthly-level articles like "What does the spectral theory say?" to popular pieces ("Mathematics as a creative art" and "The thrills of abstraction"). LAS

History, T*(14-16: 1), S*, L*. An Introduction to the History of Mathematics, Fifth Edition. Howard Eves. Saunders Coll Pub, 1983, xviii + 593 pp, $34.95. [ISBN: 0-03-062064-3] Additions include sections on the metric system, the "new math" and on Bourbaki. Expanded treatment of more recent developments in mathematics. More anecdotes, human interest stories and biographical data on later mathematicians. More figures and illustrations. Expanded Problem Studies. Lists of essay topics for student papers have been added. Convenient mathematical periods chart on inside front cover. Extended chronological table in text. Numerous particular and general references. (TR, Third Edition, October 1969; Fourth Edition, August-September 1976.) JK

Foundations, T(15-16: 1), S, L. The Logical Foundations of Mathematics. William S. Hatcher. Pergamon Pr, 1982, x + 320 pp, $38. [ISBN: 0-08-025800-X] Text covering major nonconstructive foundational systems of the last century including Frege's system, type theory, ZF set theory, Hilbert's program, Gödel's incompleteness theorem, systems of Quine, and categorical algebra. Presents basic ideas of each system plus criticism and comparison of systems. Expanded, revised, updated version of author's 1968 book (TR, January 1970; Extended Review, March 1974). KS

Foundations, S. Der Zahlbegriff. Benno Artmann. Vandenhoeck & Ruprecht, 1983, viii + 265 pp, DM 34 (P). [ISBN: 3-525-40544-8] A modern and comprehensive treatment of the number concept. Treats real and complex numbers, quaternions, infinite cardinals, nonstandard numbers. Ends with Pontrjagin's topological characterization of R, C and quaternions. JD-B

Combinatorics, P. Algebraic and Geometric Combinatorics. Ed: Eric Mendelsohn. Math. Stud., No. 65. Elsevier Sci Pub, 1982, xiii + 378 pp, $88.50 (P). [ISBN: 0-444-86365-6] Collection of papers dedicated to Nathan Mendelsohn on his 65th birthday. Note price. JRG

Number Theory, T(15), S, L. Introduction to the Theory of Numbers. Harold N. Shapiro. Wiley, 1983, xii + 459 pp, $39.95. [ISBN: 0-471-86737-3] An interesting text appropriate for seniors and beginning graduate students. Topics include divisibility, unique factorization, arithmetic functions, congruences, quadratic residues, combinatorial considerations, and the prime number theorem. Contains many exercises, references, and bibliographical notes. SG

Linear Algebra, T(14: 1), S, L. Linear Algebra with Applications to Differential Equations. P.G. Kumpel, J.A. Thorpe. Saunders Coll Pub, 1983, xii + 353 pp, $23.95. [ISBN: 0-03-060556-3] Unusual introductory text which uses linear differential and algebraic equations to motivate and illustrate topics. Greater emphasis on function spaces than normal. Optional final chapter covers power series solutions, Laplace transform. Good range of computational and theoretical exercises. KS

Linear Algebra, T(14-15: 1), L. Linear Algebra with Applications, Second Edition. Jeanne Agnew, Robert C. Knapp. Brooks/Cole Pub, 1983, xiii + 361 pp, $22.95. [ISBN: 0-534-01364-3] Written (apparently successfully) to be accessible to sophomores with limited backgrounds, the text begins with matrix algebra, and works through linear systems, determinants, and eigenvalues before developing the formalism of linear spaces, transformations, and quadratic forms. Each chapter has at least one significant application included in the text and the last chapter is devoted to linear programming. Several computer programs, in Basic, are included. (TR, First Edition, January 1979.) JS

Linear Algebra, T(13-14: 1), S. Linear Algebra with Computer Applications. Ronald I. Rothenberg. Wiley, 1983, x + 387 pp, $12.95 (P). [ISBN: 0-471-09652-0] Self-instructional format. Covers standard linear algebra topics, plus applications (Markov processes, population growth) and use of computers and programmable calculators to do matrix calculations. Does not assume calculus. JRG

Finite Mathematics, T(13-14: 1, 2). Finite Mathematics with Applications to Business, Life Sciences, and Social Sciences. Stanley I. Grossman. Wadsworth Pub, 1983, xiv + 512 pp. [ISBN: 0-534-01246-9] Topics include introductions to linear equations and matrices, linear programming, probability, statistics, game theory, and business topics. Tells students the what and the why (not just the how) in an engaging style. Wide variety of applications. A promising text. GHM

Calculus, T(13-14: 1, 2). Elements of Calculus. G. Don Allen, Charles Chui, Bill Perry. Brooke/Cole Pub, 1983, xiii + 460 pp, $24.95. [ISBN: 0-534-01188-8] Applied calculus for business, social and biological science students. Single and multivariable calculus, some differential equations and optimization. Contains many problems, index of applications. RM

Calculus, T*(13-14: 1, 2), S. Mathematics for the Biosciences. Michael R. Cullen. Prindle, Weber & Schmidt, 1983, xvii + 751 pp. [ISBN: 0-87150-352-2] Emphasis on calculus and its applications. Coverage includes vectors and matrices, discrete probability, differential and difference equations. For freshmen and sophomores with high school algebra background. Intuitive rather than rigorous treatment. Wealth of prerequisite-free applications drawn from journals and more advanced texts. Adaptable to a variety of quarter, semester, and two-semester courses. Nicely done, especially the applications in the biosciences. JK

Calculus, T(13: 2). Differential and Integral Calculus. Ya. S. Bugrov, S.M. Nikolsky. Transl: Leonid Levant. MIR Pub, 1982, 464 pp. Brief theoretical treatment, little attention paid to motivation. Essentially no problems or applications. JRG

Complex Analysis, S(18), P. Aspects of Contemporary Complex Analysis. Ed: D.A. Brannan, J.G. Clunie. Academic Pr, 1980, xiii + 572 pp, $108. [ISBN: 0-12-125950-1] Proceedings of a 1979 NATO Advanced Study Institute at Durham, England. In three parts: notes for 15 short courses, 19 lectures, progress reports on old and new problems. Topics, nearly all one-variable, include BMO, quasiconformal mappings, interpolation theory, approximation. PZ

Differential Equations, T*(14-15: 1, 2), L. Differential Equations and Their Applications: An Introduction to Applied Mathematics, Third Edition. M. Braun. Appl. Math. Sci., V. 15. Springer-Verlag, 1983, 546 pp, $24. [ISBN: 0-387-90806-4] Two significant changes have been made in this edition: the section dealing with singular solutions of differential equations has been revised and expanded, and a new section on bifurcation theory has been added. (First Edition, TR, December 1975, Extended Review, October 1977; Second Edition, TR, June-July 1978.) AO

Differential Equations, P.** Asymptotic Methods in Nonlinear Wave Theory. Alan Jeffrey, Takuji Kawahara. Pitman Pub, 1982, x + 256 pp, $39.95. [ISBN: 0-273-08509-3] Identifies concepts which form the basis of perturbation techniques currently in use to obtain an asymptotic description of nonlinear dispersive waves. Emphasis on formal development rather than theory. Unifying theme is application of various techniques to a single equation--the Boussinesq equation, a simple but typical and nontrivial nonlinear dispersive equation. Very extensive bibliography. Of special interest to researchers in applied mathematics, physics and geophysics. JK

Numerical Analysis, P. Lecture Notes in Mathematics-973: Matrix Pencils. Ed: B. Kågström, A. Ruhe. Springer-Verlag, 1983, xi + 293 pp, $17 (P). [ISBN: 0-387-11983-3] A selection of contributions to a March 1982 conference sponsored by the University of Umea at Pite Havsbad, Sweden. Major focus is on algorithms for numerical analysis of the matrix pencil $A - \lambda B$. LAS

Functional Analysis, P. Transformations Intégrales et Calcul Opérationnel. V. Ditkine, A. Proudnikov. MIR, 1982, 435 pp. Chapters on Fourier, Laplace, Bessel and other transforms, and on the operational calculus, together with an extensive table of transforms. JD-B

Analysis, P. Handbook of Integral Transforms of Higher Transcendental Functions: Theory and Algorithmic Tables. O.I. Marichev. Transl: L.W. Longdon. Ser. in Math. & Its Appl. Ellis Horwood, 1983, 336 pp, $79.95. [ISBN: 0-85312-538-7] Modern theory of hypergeometric functions. Useful reference for evaluating a variety of definite integrals, with emphasis on the use of Mellin transforms. Systematic account including relevant background information. Requires some knowledge of analysis, including basics of complex variables. JK

Algebraic Geometry, P. Tata Lectures on Theta I. David Mumford. Progress in Math., V. 28. Birkhauser Boston, 1983, xiii + 235 pp, $20. [ISBN: 3-7643-3109-7] The first half of a survey of the theory of theta functions. The author describes motivating problems and classical applications, the heat equation, projective embeddings of torii, modular forms, sums of squares, and the zeta function. He concludes with an introduction to theta functions of several variables. SG

Algebraic Geometry, P. Topics in the Theory of Algebraic Groups. James B. Carrell, et al. Notre Dame Math. Lect., No. 10. U of Notre Dame Pr, 1982, ix + 181 pp, $9.95 (P). [ISBN: 0-268-01843-X] A series of five survey lectures (by Carrell, Curtis, Humphreys, Parshall, and Weisfeiler) on aspects of the theory of algebraic groups. SG

Differential Geometry, P. The Method of Iterated Tangents with Applications in Local Riemannian Geometry. J. Enrico White. Pitman Pub, 1982, xx + 252 pp, $43.95. [ISBN: 0-273-08515-8] A study of Riemannian manifolds from the point of view of iterated tangent bundles, a generalization of tensor analysis. The author develops the theory and presents applications to Gaussian and Riemannian curvature. SG

Operations Research, P. Mathematical Models of Renewable Resources. Ed: Roland Lamberson. Humboldt State University (Mathematical Modeling Group), 1982, 107 pp, $5 (P). Proceedings of First Pacific Coast Conference on Mathematical Models of Renewable Resources, April 1982. JRG

Probability, P*. Extremes and Related Properties of Random Sequences and Processes. M.R. Leadbetter, Georg Lindgren, Holger Rootzén. Ser. in Stat. Springer-Verlag, 1983, xii + 336 pp, $36. [ISBN: 0-387-90731-9] Divided into four parts: classical extreme value theory (dealing with sequences of independent, identically distributed random variables), extensions to dependent sequences, extensions to continuous parameter stationary processes, and applications. Good set of references. RSK

Statistics, P. The Jackknife, the Bootstrap and Other Resampling Plans. Bradley Efron. SIAM, 1982, vii + 92 pp, $12.50 (P). [ISBN: 0-89871-179-7] Lectures given at an NSF-sponsored CBMS Regional Conference at Bowling Green State University in June, 1980. Concerned with nonparametric estimation of bias, variance, and more general measures of error, using these relatively new techniques (which are simple but require extensive computations). RSK

Statistics, T(15-17: 1), P*. Sampling for Health Professionals. Paul S. Levy, Stanley Lemeshow. Lifetime Learning Pub, 1980, xv + 320 pp, $27. [ISBN: 0-534-97986-6] Elementary introduction to sample survey methodology, designed primarily for practicing statisticians in health-related fields, particularly those with minimal formal training in statistics. Thorough treatment of simple random, systematic, stratified and cluster sampling. RSK

Statistics, S(15-17), P. The Beginning Forecaster: The Forecasting Process Through Data Analysis. Hans Levenbach, James P. Cleary. Lifetime Learning Pub, 1981, xiii + 372 pp, $31.50. [ISBN: 0-534-97975-0] First of a two-volume work (see TR of The Professional Forecaster below), this book describes a number of basic forecasting methods. Divided into three parts: approaching a forecasting problem, solving a forecasting problem, and managing the forecasting function. RSK

Statistics, S(16-18), P. The Professional Forecaster: The Forecasting Process Through Data Analysis. James P. Cleary, Hans Levenbach. Lifetime Learning Pub, 1982, xiii + 402 pp, $31.50. [ISBN: 0-534-97960-2] Describes up-to-date methods for the experienced forecaster. Robust/resistant methods, regression analysis, explanatory data analysis, the Box-Jenkins methods for time series, analysis of residues. FLW

Statistics, T(17: 1), P. Analyzing Experimental Data by Regression. David M. Allen, Foster B. Cady. Lifetime Learning Pub, 1982, xvi + 394 pp, $29.95. [ISBN: 0-534-97963-7] Unified approach to classical regression, experimental design, analysis of covariance and some advanced topics, using regression techniques with appropriate design matrices. Computational algorithms are emphasized and amply illustrated on small data sets. Analyzes many real data sets. RSK

Statistics, P. Lecture Notes in Statistics-16: Specifying Statistical Models: From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches. Ed: J.P. Florens, et al. Springer-Verlag, 1983, xii + 204 pp, $14 (P). [ISBN: 0-387-90809-9] Twelve papers presented at the Second Franco-Belgian Meeting of Statisticians held in Louvain-la-Neuve (Belgium) in October, 1981. Classified according to four types of methodological considerations: specification, model approximation, robustness, and adaptivity. RSK

Statistics, P. Nonlinear Renewal Theory in Sequential Analysis. Michael Woodroofe. SIAM, 1982, v + 119 pp, $14.50 (P). [ISBN: 0-89871-180-0] Based on lectures given at an NSF-supported CBMS Regional Conference at Oklahoma State University in July, 1980. Deals with repeated significance tests, developing the mathematical techniques for determining their properties. RSK

Statistics, T(13-16: 1, 2). Introduction to Statistical Thinking. E.A. Maxwell. Prentice-Hall, 1983, xii + 574 pp, $25.95. [ISBN: 0-13-498105-7] Traditional topics for a general introduction to statistical methods for students without a strong mathematics background (no calculus is used). Emphasis on inferential statistics. Examples from a wide range of disciplines. GHM

Statistics, S(18), P. The Advanced Theory of Statistics, Volume 3: Design and Analysis, and Time Series, Fourth Edition. Sir Maurice Kendall, Alan Stuart, J. Keith Ord. Macmillan Pub, 1983, x + 780 pp, $65. [ISBN: 0-02-847860-6] Revision of the final volume of this classic treatise, which deals with the analysis of variance, design of experiments, sample survey theory, multivariate analysis, and time-series. Major changes are in the latter two areas, reflecting recent developments. RSK

Statistics, S(13-18), P, L. Statistics in Medical Research: Methods and Issues, with Applications in Cancer Research. Ed: Valerie Miké, Kenneth E. Stanley. Wiley, 1982, xxi + 551 pp, $34.95. [ISBN: 0-471-86911-2] Papers from a 1981 conference on biostatistics in clinical oncology. Provides an overview of biostatistical methods as well as practical aspects of medical research. FLW

Statistics, T(13-14: 1). A First Course in Statistics. James T. McClave, Frank H. Dietrich, II. Dellen Pub, xii + 433 pp, $24.95. [ISBN: 0-89517-050-7] Presupposes no college mathematics. The usual topics presented in a straightforward manner. Includes some interesting case studies. FLW

Statistics, P. Lecture Notes in Medical Informatics-20: Manual for the Planning and Implementation of Therapeutic Studies. Sibylle Biefang, Wolfgang Köpcke, Martin A. Schreiber. Springer-Verlag, 1983, 100 pp, $12 (P). [ISBN: 0-387-11979-5] Translation of a 1979 German guidebook that compiled from a variety of sources, strategies and recommendations for the effective conduct of clinical studies as a prerequisite to statistical studies. "Applied statistical methods still appear to many doctors to constitute an estrangement of medicine in thought and action." LAS

Statistics, T(15-18: 1, 2), S, P, L. Applied Linear Statistical Methods. Donald F. Morrison. Prentice-Hall, 1983, xiv + 562 pp, $38.95. [ISBN: 0-13-041020-9] Presupposes a basic statistics course. Matrix algebra is presented in an appendix. Regression analysis, time series, stepwise regression, analysis of variance and covariance. Underlying assumptions emphasized but proofs are usually omitted. Real data used in examples. FLW

Statistics, T(16-18: 1), S, L. A Course in Linear Models. Anant M. Kshirsagar. Statistics, V. 45. Dekker, 1983, xvi + 422 pp, $32.75. [ISBN: 0-8247-1585-3] Presupposes matrix theory and a course in mathematical statistics. Supplies the theory behind regression analysis, analysis of variance and covariance, generalized least squares, and variance components analysis. FLW

Statistics, T(16-18: 1, 2), S, P, L. An Introduction to Applied Multivariate Statistics. M.S. Srivastava, E.M. Carter. Elsevier North-Holland, 1983, x + 394 pp, $34.50. [ISBN: 0-444-00621-4] Presupposes some statistics. Supplies a chapter on matrix algebra. Multivariate normal distributions, analysis of variance, and regression. Growth curves. Profile, discriminant, principal component, and factor analysis. Inference on covariance matrices. FLW

Statistics, S(15-18), P. Reliability in the Acquisitions Process. Ed: Douglas J. DePriest, Robert L. Launer. Lecture Notes in Stat., V. 4. Dekker, 1983, viii + 199 pp, $35 (P). [ISBN: 0-8247-1792-9] Papers from a 1981 workshop held to "review recent developments in reliability research and to provide a forum to facilitate communications between DOD personnel and academic researchers." Note the high price for a paperback. FLW

Computer Programming, T*(13-14: 1), S*, L. Oh! Pascal! Doug Cooper, Michael Clancy. WW Norton, 1982, xix + 476 pp, $15.95 (P). [ISBN: 0-393-95205-3] Detailed and readable introduction to programming and Pascal. Includes comments on problem-solving techniques. Section on debugging, self-test and exercises at end of each chapter. Appropriate for self-study or classroom use and with batch or interactive system. KS

Computer Programming, P. Programming the TI-59 & the HP-41 Calculators. Paul Garrison. TAB Books, 1982, vi + 294 pp, $12.95 (P). [ISBN: 0-8306-1442-7] A very detailed introduction to the TI-59 and HP-41 for the complete calculator novice. A great number of examples. But what novice will have the patience to learn these two quite different programming systems simultaneously before choosing one or the other? GHM

Computer Programming, S*(13-16), P*, L.** The Programmer's Book of Rules. George Ledin, Jr., Victor Ledin. Lifetime Learning Pub, 1979, 248 pp, $12.95 (P). [ISBN: 0-534-97993-9] Two-hundred fifty Do's and Don'ts, with brief explanations and examples. These "state-of-the-art" maxims are supported by numerous references to recent literature on programming methodology. GHM

Computer Programming, P, L. The Preparation of Programs for an Electronic Digital Computer, with special reference to the EDSAC and the use of a library of subroutines. Maurice V. Wilkes, David J. Wheeler, Stanley Gill. Tomash Pub, 1982, xxxi + 170 pp, $30. [ISBN: 0-938228-03-X] Photographic reproduction of the "first programming textbook" ever published (1951). Introduction by Martin Campbell-Kelly outlines the origins, innovations and influence of this detailed guide to the celebrated EDSAC computing machine which became operational at Cambridge in 1949. GHM

Computer Programming, T(13: 1), S. Programming in BASIC. Tom Logsdon. Anaheim Pub, 1981, ix + 258 pp, $15.95 (P) [ISBN: 0-88236-179-1]; Instructor's Guide and Answer Manual, 101 pp, (P). Slow-paced introduction to BASIC. Covers only one-dimensional arrays. Exercises and programming problems at end of each chapter. Assumes student is using a terminal. Some remarks, pictures and language may be regarded as sexist. Instructor's Guide includes listings and runs for programming problems. KS

Computer Programming, T(13-14: 1). Structured Fortran 77 Programming. Seymour V. Pollack. Boyd & Fraser Pub, 1982, xvi + 496 pp, $17.95 (P). [ISBN: 0-87835-095-0] A structured introduction to the programming language Fortran 77. The author spends the first two chapters introducing the concepts of an algorithm and program design. The book then introduces the complete language specifications in Chapters 3-17. The examples all appear to be well structured, within the limits of what is available on Fortran 77. The order of presentation seems quite reasonable. This book seems comparable to a number of good Fortran 77 books on the marketplace. MS

Computer Programming, P. Z80 Assembly Language Subroutines. Lance A. Leventhal, Winthrop Saville. Osborne/McGraw-Hill, 1983, xi + 497 pp, $15.95 (P). [ISBN: 0-931988-91-8] This is a text for experienced assembly language programmers who must learn the specific details of the assembly language of the Z80 processor. It contains only a one-chapter introduction to the concepts of assembly level programs while Chapters 2-11 are a description of the features available on the Z80. In addition, it contains many complete subroutines for common system level operators such as code conversion, bit manipulation, string manipulation, and interrupt processing. This is mainly an advanced reference book rather than a text. MS

Computer Programming, T(14: 1), S. The Programmer's Craft: Program Construction, Computer Architecture, and Data Management. Richard J. Weiland, Charles R. Bauer. Reston Pub, 1983, xv + 160 pp. [ISBN: 0-8359-5645-8] A mix of structured programming (language independent), computer architecture, and data structures, with a data processing orientation. Probably too sparse for a text. RM

Software Systems, S(15-16), P, L. Language Translators. John Zarrella. Microprocesor Software Eng. Concepts Ser. Microcomputer Applic, 1982, ix+ 189 pp, $12.95 (P). [ISBN: 0-935230-06-8] Not geared to any particular language or machine, this book gives a fine overview of the general principles of assemblers, compilers and interpreters for anyone interested in how programs are converted into machine-executable form. Presumes a slight familiarity with machine and assembly language programming. GHM

Software Systems, S(16-18), P. Distributed Systems--Architecture and Implementation, An Advanced Course. Ed: B.W. Lampson, M. Paul, H.J. Siegert. Springer-Verlag, 1983, xiii + 510 pp, $17 (P). [ISBN: 0-387-10571-9] Originally published as Volume 105 in the Lecture Notes in Computer Science series. Contains the texts of twenty papers presented at the Technische Universität München, March 4-13, 1980. The papers cover various aspects of the design and implementation of distributed systems. AO

Computer Science, T(13: 1). Mathematics for Data Processing, Second Edition. Frank J. Clark. Reston Pub, 1983, x + 326 pp, $21.95. [ISBN: 0-8359-4263-5] Intended for students seeking entry level occupations in data processing. Covers numeration systems, Boolean algebra, functions, matrix algebra, introductory material in linear programming and statistics. (First Edition, TR, January 1975.) JRG

Computer Science, P. Lecture Notes in Computer Science-148: Logics of Programs and Their Applications. Ed: A. Salwicki. Springer-Verlag, 1983, vi + 324 pp, $14.50 (P). [ISBN: 0-387-11981-7] 23 selected contributions to an August 1980 conference in Poznań, Poland dealing with program logics (including methods from nonstandard logic and model theory), abstract data structures, parallel computations and the University of Warsaw's programming language LOGLAN. LAS

Computer Science, T(14-15: 1). The Logic Design of Computers, An Introduction. M. Paul Chinitz. Howard W. Sams, 1981, 413 pp, $15.95 (P). [ISBN: 0-672-21800-3] An introduction to the principles of computer organization and logic design intended for students who have had an introduction to programming languages and are now interested in looking at how these ideas are actually carried out

within a computer system. It looks at both the functional organization and the logic design level of a typical digital computer. MS

Computer Science, P. Third Caltech Conference on Very Large Scale Integration. Ed: Randal Bryant. Computer Sci Pr, 1983, xii + 430 pp, $36.95. [ISBN: 0-914894-86-2] This is a collection of 23 papers, all of which were presented at the Third Conference on VLSI, Pasadena, California, March 1983. The papers review what has occurred in the area of design, verification, and testing of VLSI circuits over the past ten years. The papers are quite technical and theoretical in nature and would probably be appropriate for graduate students in electrical engineering, physics, or computer science. MS

Computer Science, T(16-17: 1), L. Principles of Programming Languages: Design, Evaluation, and Implementation. Bruce J. MacLennan. Holt, Rinehart & Winston, 1983, xvi + 544 pp. [ISBN: 0-03-061711-1] Principles of programming language design. Organization treats important languages (Fortran, Algol, Pascal, Ada, Lisp, Smalltalk, Prolog) as wholes to emphasize interrelationships between the parts, in addition to comparisons between the languages. Emphasis on historical context, as well as special features of the languages. RM

Control Theory, P. Lecture Notes in Control and Information Sciences-47: Adaptive Systems with Reduced Models. P.A. Ioannou, P.V. Kokotovic. Springer-Verlag, 1983, v + 164 pp, $10 (P). [ISBN: 0-387-12150-1] A research monograph on the robustness properties of model reference schemes (a class of adaptive systems). AO

Applications (Biology), S(15-17), P. An Introduction to Mathematical Taxonomy. G. Dunn, B.S. Everitt. Stud. in Math. Biology, V. 5. Cambridge U Pr, 1982, xi+ 152 pp, $29.95; $11.95 (P). [ISBN: 0-521-23979-6; 0-521-28388-4] Philosophy and methodology behind mathematical techniques of classification in biology, including principal components analysis, multidimensional scaling, and cluster analysis. Requires only minimal background in elementary statistics and matrix algebra. Good set of references. RSK

Applications (Health and Public Affairs), T(17: 1), P*. Public Program Analysis: A New Categorical Data Approach. Ron N. Forthofer, Robert G. Lehnen. Lifetime Learning Pub, 1981, xx + 294 pp, $30.95. [ISBN: 0-534-97974-2] Describes and illustrates the weighted least squares approach for analyzing categorical data in the areas of public health and public affairs. Includes both single and multiple-response applications. Presumes background in matrix algebra and analysis of variance. RSK

Applications (Information Science), P. Lecture Notes in Computer Science-146: Research and Development in Information Retrieval. Ed: Gerard Salton, Hans-Jochen Schneider. Springer-Verlag, 1983, ix + 311 pp, $14.50 (P). [ISBN: 0-387-11978-7] Proceedings of a May 1982 conference in Berlin of the ACM's special interest group on information retrieval (SIGIR). A remarkably broad interdisciplinary study, linking library science, computer science and linguistics. Sample result: text compression in English using prefixes and suffices is NP-hard. LAS

Applications (Physics), P. Geometric Quantization in Action: Applications of Harmonic Analysis in Quantum Statistical Mechanics and Quantum Field Theory. Norman E. Hurt. D Reidel Pub, 1983, xiv + 336 pp, $49.50. [ISBN: 90-277-1426-6] Quantum statistical mechanics and quantum field theory are used to illustrate the use of geometric quantization techniques. AO

Applications (Physics), T(17-18: 1, 2), P. Mathematical Foundations of Elasticity. Jerrold E. Marsden, Thomas J.R. Hughes. Prentice-Hall, 1983, xviii + 566 pp, $42.95. [ISBN: 0-13-561076-1] A well-written introduction to the mathematical theory of three-dimensional elasticity using modern differential geometry and functional analysis. AO

Applications (Simulation), T(15-16: 1). Introduction to Simulation: Programming Techniques and Methods of Analysis. James A. Payne. McGraw-Hill, 1982, xii + 324 pp, $33.95. [ISBN: 0-07-048945-9] A survey of techniques for performing simulation on a digital computer. It assumes that the reader has a knowledge of programming in a high-level language as well as a fundamental background in statistics. It concentrates most heavily on the techniques of discrete-event simulation and introduces the special purpose languages SIMSCRIPT and GPSS for implementing discrete event simulation models. MS

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Southeastern Section

The sixty-second annual spring meeting of the Southeastern Section met at The Citadel, Charleston, South Carolina on April 15-16, 1983. A total of 255 persons attended the meeting.

Invited Lectures:

"Recent Progress on Combinatorial Problems in Ramsey Theory and Discrete Geometry," by William T. Trotter, Jr., University of South Carolina.
"Beyond the Binomial," by John Neff, Georgia Institute of Technology.
"Paradoxes About Rationals and Irrationals," by Ivan Niven, University of Oregon.

Special Sessions:

"Domination Number in Graphs," by Michael Johnson, University of Louisville.
"Strongly Regular Graphs," by Robert Roth, Emory University.
"Cycles in Hypergraphs," by Richard Duke, Georgia Institute of Technology.
"Hamiltonian Properties in Graphs," by Ronald Gould, Emory University.
"On Hamiltonian Cycles in Cayley Color Graphs," by Joseph B. Klerlein, Western Carolina University.
"Algorithms for Shortest Paths," by Douglas R. Shier, Clemson University.
"Neighborhood Relations in Finite Graphs," by David Sumner, University of South Carolina.
"Some Algebras Associated with Graphs," by Trevor Evans, Emory University.

Short Presentations:

"On Invertible Linear Combinations of Matrices," by Peter M. Gibson, University of Alabama in Huntsville.
"Shuffling Cards, Imprimitivity, and Correcting Errors," by Stephen L. Davis, Davidson College.
"On the Computer Algorithm of Generalized Euclidean Algorithm in the Ring of Integral Matrices," by Jau-shyong Shiue, Gardner Webb College.
"Infinitesimal Deformations of Coordinate Axes," by Charles G. Fleming, The Citadel.
"Affine G-Mappings," by Irl C. Bivens, Davidson College.
"An Algorithm for Monitoring Global Error in Numerical Solution of a System of Linear Differential Equations," by James C. Pleasant, East Tennessee State University.
"An Ad Hoc Method for Identifying Absolute Extrema in Lagrange Multiplier Problems," by John Baxley, Wake Forest University.
"Transformations of Circles in E_2 as Point Transformations in E_3," by Ray Wylie, Furman University.
"On Convergence and Regularity of Regular Ritt Series," by G.R. Viswanath, South Carolina State College.
"A Factorization Theorem for an $H^2(\mathcal{V})$ Space," by Nancy Lee Shell, Furman University.
"Teaching Remedial Math in College: Is It Possible? A Survey of Developmental Mathematics in College," by Ping-Tung Chang, Augusta College.
"The Arithmetic-Geometric Mean Inequality: Optimization Without Calculus," by Larry Riddle, Emory University.
"The Sloan Program in the New Liberal Arts at Davidson," by R. Bruce Jackson, Jr., Davidson College.
"Optimizing Boxes and Disks," by Thomas A. Hern, University of North Carolina at Chapel Hill/Bowling Green State University.
"A Monte Carlo Method for Integration in Polar Coordinates," by Jerry E. Bolic, Lenoir-Rhyne College.
"Some Elementary Calculus Examples from Statistics," by Richard G. Vinson, University of South Alabama.
"A Matter of Life and Death: A Markovian Process," by Subhash C. Saxena, Coastal Carolina College.
"Carroll's Obtuse Triangle Problem," by Douglas H. Frank, Valdosta, Georgia.
"The Use of the Stem and Leaf Plot in Constructing a Frequency Table," by Lloyd B. Smith, Jr., Lenoir-Rhyne College.
"Modern Cryptography: The Mathematics of Secrecy," by Michael Willett, University of North Carolina at Greensboro.
"Teaching Mathematical Induction Via Computing," by Paul R. Patten, North Georgia College.
"Using the Microcomputer to Discover Patterns in the Graph of Polar Equations," by Linda H. Boyd and Charles R. Stone, Dekalb Community College.
"A Study of Students Who Took Basic Math, Fall 1980," by Sandria N. Kerr, Winston-Salem State University.
"Representation of a Set Whose Derived Set is Countable," by Arthur G. Sparks, Georgia Southern College.
"Alternative Rings with Commuting Nilpotent Elements," by Tae-il Suh, East Tennessee State

University.
"A Lagrange Multiplier Problem with Multiple Critical Points," by Ted Gentry, Wake Forest University.
"Some Relations Between Groups and Graphs," by Terri England, Emory University.
"A Computer Study of Amniocentesis Data," by Cynthia Priest, Emory University.
"Implementation of a Spelling Checker and Corrector on a Microcomputer," by Kevin Pepe, Emory University.
"Almost Periodic Functions and Semigroups," by Arnold Goldstein, Savannah, Georgia.
"Weighted Associated Stirling Numbers," by Fred T. Howard, Wake Forest University.
"Some Variations on the Problem of Generating Pythagorean Triples," by Kenneth E. Whipple, Georgia State University.
"Point Unstable Graphs," by Ted R. Monroe, Converse College.
"Color Algebras," by Steven D. Comer, The Citadel.
"Which Triangular Numbers are Square?" by David R. Stone, Georgia Southern College.
"Schedules for Testing Individuals in Tournaments for Pairs," by Gerald Huff, University of Georgia.
"Weak Openness and Almost Openness," by David A. Rose, Francis Marion College.
"A Division Game or How Far Can You Trust Mathematical Induction," by William H. Ruckle, Clemson University.
"Polygons in Ordered Planes," by R.B. Killgrove, University of South Carolina at Aiken.
"Casting Shadows on Eggs," by Robert E. Jamison, Clemson University.

Florida Section

The sixteenth annual spring meeting of the Florida Section was held on March 4-5, 1983 at Florida State University, Tallahassee. There were 128 registrants.

Invited Addresses:

"Pure and Applied Topology in 2, 3, and 4-Space," by T. Benny Rushing, University of Utah and Institute for Advanced Study.
* "The Mathematical Sciences K-12: What Is Still Fundamental and What Is Not," by Marcia P. Sward, Associate Director, MAA.
"Signs Sum Square Summable Sequences," by James R. Retherford, Louisiana State University.
"Mathematicians in Operations Research in the U.S. Army Air Force in World War II," by Charles W. McArthur, Florida State University.
"Motivation in Mathematics: Believe It or Not, It Is a Record," by Ignacio D. Bello, Hillsborough Community College.
"A Math-Science Program for Gifted Girls," by Paul E. McDougle, University of Miami.
"Antoine's Necklace: Or How to Keep a Necklace from Falling Apart," by Beverly L. Brechner, University of Florida.

Short Presentations:

* "A Dialogical Theory of Mathematics Education," by James W. Garrison, Florida State University.
"Linear Algebra and Space Travel to Music on the Apple Microcomputer," by Gareth Williams, Stetson University.
"Factoring Binomials and Trinomials in Community Colleges with Questions for the 'Gordon Rule'," by Carlton A. Lane, Hillsborough Community College.
"Harmonizing with a Calculator," by Alan Wayne, Pasco-Hernando Community College.
"Eigenvalues and Eigenvectors of 2x2 Matrices," by James R. Weaver, University of West Florida.
"A Computer Simulation of the Unlimited Register Machine," by Marion G. Harmon, Florida State University.
"A Survey of Pending Florida and Federal Programs for Mathematics Education," by E.P. Miles, Florida State University.
"Open Mappings and Dimension," by Alice Mason, University of Florida.
"Applying Bessel Functions Underground," by John D. Hall, Southwest Florida Water Management District.
"Partial Fractions," by Shiv K. Aggarwal, Embry-Riddle Aeronautical University.
"Report on Remedial/Developmental Mathematics Conference," by Donald M. Hill, Florida A&M University.

Panel Discussion:

"CUPM Curriculum Recommendations," by Gareth Williams (Moderator), Stetson University; Bettye Anne Case, Abraham Kandel, and Frederick Leysieffer, Florida State University; Kermit N. Sigmon, University of Florida.
"Emmy Noether, Her Life and Work," organized by Association for Women in Mathematics: "Noether's Life," by Betsey Whitman, Florida A & M University; "Noether's Work in Abstract Algebra," by Robert Gilmer, Florida State University.
"MAA Placement Test Program," by Roy C. Jones, Jr. (Moderator), University of Central Florida; Gloria Child, Rollins College; Art Crummer, University of Florida; Donald M. Hill, Florida A&M University; Linda W. Smith, Tallahassee Community College.

Student Papers:

"Mathematical Concepts Represented Through Color Graphics," by Douglas R. Martin, Florida State University.
"A Mathematical Model of the Spread of Epidemic Diseases," by Robert S. Knego, University of South Florida.
"A Low Cost Design for a Color Sensing Device for the Blind," by Raymond Curci, Florida State University.
"Remedial/Developmental Mathematics Conference," organized by Donald M. Hill, Florida A & M University.

At the business meeting an award was presented to Herman Meyer of the University of Miami, Barry University, and Florida International University for outstanding contributions to mathematics and the mathematical community of the Florida Section.

Southwestern Section

The Southwestern Section held its spring meeting on March 25-26, 1983 at New Mexico Institute of Mining and Technology in Socorro, New Mexico. A total of 44 members were registered for the meeting. Bonita Ross, a consultant for MATHFILE gave a presentation and demonstration of the database MATHFILE.

Invited Address:

"Elementary Versions of Some Advanced Theorems," by Mark Kac, University of Southern California.

Short Presentations:

"The Dirichlet Problem with Unbounded Data," by David Siegel, New Mexico Institute of Mining and Technology.
"The Rapid Computation of Residues," by Arthur Knoebel, New Mexico State University.
"Applications of Grassman Algebra to Graphical Statistics," by Alvin Swimmer, Arizona State University.
"The Second Fundamental Problem in the Theory of Quadratic Residues," by Justin McCarthy, Deming, New Mexico.
"The Uniqueness of Generating Functions for Shifted Tableau, or The Hashing of Multisets," by Timothy Kraus, University of New Mexico.
"Computers and Pattern Identification," by Peter Bonner, New Mexico Highlands University.
"Non-additive Measures and Random Sets," by Hung T. Ngyen, New Mexico State University.
"On the Distributions of LRT for Testing the Homogeneity of Several Exponential Populations," by Anita Singh, New Mexico Institute of Mining and Technology.
"Empirical Bayesian Estimation in Life Testing," by Ashok K. Singh, New Mexico Institute of Mining and Technology.
"Decision Making in Baseball Using Simulation," by Stephen Thorpe, New Mexico Institute of Mining and Technology.
"Isospectral Potentials for the Laplacian," by H.D. Fegan, University of New Mexico.
"Some Remarks on the Free Surface Condition for the Two-dimensional Laplace Equation," by Ross Lomanitz, New Mexico Institute of Mining and Technology.
"Periodic Solutions of Autonomous Matrix Riccati Differential Equations," by David Sanchez, University of New Mexico.
"A Modified Harminic Series," by Clyde M. Dubbs, New Mexico Institute of Mining and Technology.
"Elementary Proofs for Some Advanced Theorems," by Piotr Antosik, UTEP.
"Statistics About the Mathematical Preparation of Entering Freshmen at UTEP, 1964-1983," by J.R. Provencio, UTEP.

Ohio Section

The annual spring meeting of the Ohio Section was held April 22-23, 1983.

Invited Addresses:

"Mathematical Modeling: When, Why, and For Whom," by Maynard Thompson, Indiana University.
"Topology, Proofs, and CUPM's Mathematical Sciences Program Recommendations," by David Lutzer, Miami University.
"A Mathematics Department Microcomputer Laboratory," by Maynard Thompson, Indiana University.
"The Contributions of Anna Johnson Pell Wheeler," by Nazra Azarnia, Miami-Hamilton.
"How to Keep Students Awake: Some Tricks of the Trade," by Darrell Horwath, John Carroll University.

Panel Discussions:

"Initial Recommendations for Teacher Certification: What Do You Think?" by Dick Shumway (Moderator), Ohio State University.
"Report on School-University Articulation," by William Beyer (Moderator), University of Akron.
"Panel on Graduate Study Options," by David Lipsich (Moderator), University of Cincinnati.
"Microcomputer Software Swap Session," by Robert Anderson (Coordinator), Marietta College.

Contributed Papers:

"Generating Fuzzy Topologies with Semi-closure Operators," by Albert J. Klein, Youngstown State University.
"Pavel Sergeevich Alexandorff: In Memoriam," by Douglas E. Cameron, University of Akron.
"Quotient Additive Properties in Topological Spaces," by Norman Levine, Ohio State University.
"Near-best Approximation by Averaging Polynomial Interpolants," by Judith Palagallo, University of Akron.
"Conic Sections-1P vs. Euclidean," by Richard Laatsch, Miami University.
"Angular Separations of Finite Sets in E^2," by E.P. Merkes, University of Cincinnati.
"Study of Curves at L_∞" by Kenneth Cummins, Kent State University.
"The 1983 AHSME, Mathematical Content and Implications of the Results," by Leo J. Schneider, John Carroll University.
"Modula-2: Usage in Modelling and Simulation," by Ralph Hollingsworth, Muskingum College.
"Promoting COBOL," by William M. Wagner, Dublin, Ohio.

Kansas Section

The sixty-eighth annual spring meeting of the Kansas Section was held on April 8-9, 1983 at the University of Kansas, Lawrence, Kansas. Approximately 125 persons attended.

Invited Addresses:

"An Anecdotal History of Mathematics in Kansas," by G. Baley Price, University of Kansas.
"Approximate Arithmetic Progressions," by Ronald L. Graham, Bell Telephone Laboratories.

Panel Discussion:

"The Mathematics Challenge in Two-Year Colleges," by Jean Moran (Moderator), Donnelly College.

Contributed Papers:

"Computer Graphics in the Teaching of Mathematics," by Earl J. Schweppe, University of Kansas.
"3-D Image Generation in Vector Calculus," by Michael Frantz, Wichita State University.
"A Class of Sequences Related to Powers of the Golden Ratio," by Hung Cao, Wichita State University.
"What are Dualities?" by Juergen Koslowski, Kansas State University.
"Approximate Roots of Polynomials," by Chris Christensen, University of Kansas.
"Prime Recognition Using a Probabilistic Algorithm on an Apple II," by Dave Foley, Wichita State University.
"Fibonacci Structure Made Visual," by Larry Schulte, University of Kansas.
"Another Look at the Power Rule," by Bonnie Ernst, Wichita State University.
"Symbolic Integration Algorithms," by William David Miller, Kansas State University.
"Incorporating Microcomputers Into the Curriculum," by Wayne F. Mackey, Johnson County Community College.
"On the Problems of Presenting Graphically the Shape of Non-Symmetric Solids of Higher Dimension," by T.M. Creese, University of Kansas.
"A Note on the Partial Differential Equation $u_{xx} + u_{yy} = f$" by William M. Self, Pittsburg State University.
"An Overview of Latin Squares and Their Orthogonality," by D.V. Chopra and Dale Hughes, Wichita State University.
"Generation of Families of Combinatorial Structures," by Thomas C. French, Wichita State University.
"Differentials of Numerical Functions," by Louis Herman, Kansas State University.
"See How They Run: Phenotyping Race Horses with Polar Coordinates," by Paul Mostert, University of Kansas.
"SIMPL: Remedial Mathematics at the University of Kansas," by Philip R. Montgomery, University of Kansas.
"Who Was James Bernoulli?" by Glenn R. Shafer, University of Kansas.
"A Characterization of Locally Compact Spaces," by Daniel Grubb, Kansas State University.
"Forty-four Theorems for Four Fours," by Lester E. Laird, Emporia State University.
"Super Heronian Triangles," by Ruth Janssen and William H. Richardson, Wichita State University.
"Remarks on Some Submanifolds of Sasakian Manifolds," by Gregory Ronsse, Kansas State University.

North Central Section

The spring meeting of the North Central Section was held April 22-23, 1983 at Carleton College with 131 persons in attendance.

Invited Addresses:

* "Strategies for Algorithm Development and Implementation in Vehicle Routing," by Kendall E. Nygard, North Dakota State University.
"Non-Commutative Limits," by Paul Halmos, Editor of the Monthly, Indiana University.

Panel Discussion:

"Placement, Etc.," by Gail Earles (Moderator), St. Cloud State University; Edwin Andersen, Southwest High School, Minneapolis; Mary B. Johnson, Inver Hills Community College; William Tomhave, University of Minnesota at Morris.

Contributed Papers:

* "The Solution of the Dumonceaux Problem," by H.B. Coonce, Mankato State University.
"Lattices of Metrics," by Harold W. Martin, St. Cloud State University.
"The Axis and Angle of Rotation Corresponding to an Orthogonal Matrix," by Warren S. Loud, University of Minnesota.
* "A Residential High School for Science and Mathematics in Minnesota," by Joseph A. Gallian, University of Minnesota at Duluth.
"Computerized Black Box Exercises to Reinforce Abstract Concepts in Calculus," by George Mills, Carleton College.
"An Example of Bayesian Estimation in Time Series," by Loretta Thielman, Gustavus Adolphus College.
* "An Heuretical Approach to a Well-Studied Problem," by Sylvan Burgstahler, University of Minnesota at Duluth.
"What is the Fast Fourier Transform?" by J. Arthur Seebach, St. Olaf College.

Michigan Section

The spring meeting of the Michigan Section was held on May 6-7, 1983 at the Oakland Community College, Auburn Hills Campus, Michigan. The meeting was attended by approximately 100 members.

Invited Addresses:

"Contemporary Applications of Linear and Integer Programming," by Harvey M. Salkin, Case Western Reserve University.
"Using Quantitative Techniques to Take the Gambling Out of the Stock Market," by Harvey M. Salkin, Case Western Reserve University.
"National Perspectives on Mathematical Education," by Marcia Sward, Associate Director, MAA.
"Finite Groups and Finite Geometry," by Jonathan I. Hall, Michigan State University.
"Applying Mathematics in Industry," by W. Weston Meyer, General Motors Research Labs.

Short Presentations:

"Geometric Methods in Babylonian Mathematics," by James K. Bidwell, Central Michigan University.
"Who Wants to Talk About Mathematics in China?" by Chester Tsai, Michigan State University.
"Using Microcomputers to Illustrate Concepts in Probability and Statistics," by Elliot A. Tanis, Hope College.
"Exploratory Data Analysis Using Microcomputers," by Tom Ten Hueve III and Lynn Ploughman, Hope College.
"A Proposed Logic Machine Control System," by Rick L. Stevens, Western Michigan University and Argonne National Lab.
"Decompositions of Graphs," by Sergio Ruiz, Western Michigan University.
"The Euclidean Algorithm Revisited," by John W. Petro, Western Michigan University.
"Midwest Talent Search," by Sharon Higham (Coordinator), Michigan Talent Search.
"Finding All Solutions to a System of Polynomial Equations," by Alden H. Wright, Western Michigan University.
"Using the Gutenberg Word Processor," by P.K. Wong, Michigan State University.
"Computer Graphics: Surfaces in R^3," by Jack Kuipers, Calvin College.
"Public Secret Codes," by F.I. Papp, University of Michigan at Dearborn.

Northern California Section

The 1983 spring meeting of the Northern California Section was held February 26, 1983 at Stanford University. There were 215 registered participants at the meeting.

Invited Lectures:

"Imagery in Mathematics Applications," by Ross L. Finney, Massachusetts Institute of Technology.
"Some Paradoxical Coverings of the Real Line," by Ivan Niven, University of Oregon.
"Mathematical Analysis of Decorative Patterns," by Dorothy K. Washburn, California Academy of Science.
"Hilbert's 'Grundlagen der Geometrie' Revisited," by Garrett Birkhoff, Harvard University.
"Anecdotes from the Early History of Computing," by Henry S. Tropp, Humboldt State University.

Indiana Section

The spring meeting of the Indiana Section was held at Indiana University on April 19, 1983 with approximately 60 members present. The Indiana Small College Math Competition was held in conjunction with the meeting.

Invited Addresses:

"Puzzle It Out," by John Ewing, Indiana University.
"Public Key Cryptography," by Dorothy Denning, Purdue University.
* "Finding All Solutions to Small Systems of Polynomials Using a Computer," by Alexander Morgan, General Motors Research Institute.
"Algorithmically Defined Functions," by Richard D. Anderson, Louisiana State University.

Iowa Section

The 70th annual spring meeting of the Iowa Section was held on the campus of Iowa State University, Ames, Iowa on April 22-23, 1983. The meeting was held jointly with the Iowa Section of SIAM. Altogether 83 people attended the meetings.

Invited Addresses:

"Dynamics on the Riemann Sphere," by Paul Blanchard, Boston University.
"Classroom Notes," by Leonard Gillman, University of Texas.

Student Papers:

"A Ubiquitous Partition on Perfect Subsets of R^n," by Donald John Nicholson, Iowa State University.
"The Apple Looks at Bertrand's Paradox," by Todd Wibben and Jimmie Arnold, Iowa Wesleyan College.
"Factoring the Characteristic Polynomial for Graphs with Symmetry," by Grant Izmirlian, Drake University.
"Cardano's Method of Solving Cubics in Finite Fields," by Doug Jaynes and Bob Anderson, Iowa Wesleyan College.

Short Presentations:

"Applications of Clifford Algebras in Theoretical Physics," by Nikos Salingaros, University of Iowa.
"Automata: A View from Four Perspectives," by Michael H. Millar, University of Northern Iowa.
"A Geometric Limit Problem in Calculus," by Tom Whaley, Central College.
"Hyperspace--Exploring the Concept of Dimension," by George Kelley, Maharishi International University.
"A Fuller Value of Geometry," by George Kelley, Correy Vion and Vedder Wright, Maharishi International University.
"The Musical Experience of Number," by George Kelley and Thomas Stone, Maharishi International University.
"On Characterization of Isospectral Graphs," by Wayne L. Woodworth, Drake University.
"A Matrix Calculator for Classroom Use," by Leslie Hogben, Iowa State University.
"The Set of Solutions of a Semi-linear Parabolic Equation," by Paul Sacks, Iowa State University.
"Practical and Mathematical Aspects of Computed Tomography," by W.R. Madych, Iowa State University.
"Frequency Entrainment in Periodically Forced Oscillators," by Jim Murdock, Iowa State University.
"On Deriving the Characteristic Polynomial of a Graph," by Milan Randic, Drake University.
"Using an Understanding of Consciousness to Explain the Effectiveness of Mathematics in the Sciences," by Catherine Gorini Wadsworth, Maharishi International University.
"The Geometry of Coefficients of Powers of Polynomials," by Stephen J. Willson, Iowa State University.
"A Refined Platonist Philosophy of Mathematics," by Eric W. Hart, Maharishi International University.

Using Heron's Formula, the inequality to be proved will follow from

$$\text{(A)} \qquad \sqrt[4]{stuv} + \sqrt[4]{s't'u'v'} \leqslant \sqrt[4]{(s+s')(t+t')(u+u')(v+v')}$$

for positive $s, t, u, v, s', t', u', v'$. A simpler analogous inequality that might be helpful is

$$\text{(B)} \qquad \sqrt{xy} + \sqrt{x'y'} \leqslant \sqrt{(x+x')(y+y')} \quad \text{for} \quad x, y, x', y' \text{ positive.}$$

First we note that (B) follows from the Cauchy Inequality applied to the vectors $(\sqrt{x}, \sqrt{x'})$ and $(\sqrt{y}, \sqrt{y'})$ [and also follows from $(\sqrt{xy'} - \sqrt{x'y})^2 \geqslant 0$ or from the Inequality on the Means applied to xy' and $x'y$]. Using (B) with $x = \sqrt{st}$, $x' = \sqrt{s't'}$, $y = \sqrt{uv}$, $y' = \sqrt{u'v'}$ and reapplying (B) to the new right side, one has

$$\sqrt[4]{stuv} + \sqrt[4]{s't'u'v'} \leqslant \sqrt{(\sqrt{st} + \sqrt{s't'})(\sqrt{uv} + \sqrt{u'v'})}$$
$$\leqslant \sqrt{\sqrt{(s+s')(t+t')}\sqrt{(u+u')(v+v')}}\,.$$

Since here the rightmost part equals the right side of (A), we have proved (A).

Equality holds in (B) if and only if $\sqrt{x} : \sqrt{x'} = \sqrt{y} : \sqrt{y'}$ and this holds if and only if $x : x' = y : y'$. Hence equality occurs in (A) if and only if $s : t : u : v = s' : t' : u' : v'$. It follows that equality occurs in the original inequality if and only if a, b, c are proportional to a', b', c'.

PROGRESS REPORTS

EDITED BY THOMAS BANCHOFF AND RICHARD MILLMAN

It is easy to be too busy to pay attention to what anyone else is doing, but not good. All of us should know, and want to know, what has been discovered since our formal education ended, but new words, and relations between them, are growing too fast to keep up. It is possible for a person to learn of the title of a recent work and of the key words used in it and still not have the faintest idea of what the subject is.

Progress Reports is to be an almost periodic column intended to increase everyone's mathematical information about what others have been up to. Each column will report one step forward in the mathematics of our time. The purpose is to inform, more than to instruct: what is the name of the subject, what are some of the words it uses, what is a typical question, what is the answer, who found it. The emphasis will be on concrete questions and answers (theorems), and not on general contexts and techniques (theories). References will be kept minimal: usually they will include only one of the earliest papers in which the answer appears and a more recent exposition of the discovery, whenever one is easily available.

Everyone is invited to nominate subjects to be reported on and authors to prepare the reports. The ground rules are that the principal theorem should be old enough to have been published in the usual sense of that word (and not just circulated by word of mouth or in preprints); it should be of interest to more than just a few specialists; and it should be new enough to have an effect on the mathematical life of the present and near future. In practice most reports will probably be on progress achieved somewhere between 5 and 15 years ago.

MANIFOLDS WITH THE SAME SPECTRUM

RICHARD S. MILLMAN
Department of Mathematical & Computer Sciences, Michigan Technological University, Houghton, MI 49931

Many extremely important results in differential geometry concern the effect of the geometry of an object upon its topology. In addition, over the past forty years there has been a great deal of interest in the relationship between analytic quantities and the geometry and topology of

generalized surfaces. Many of these new results center on the notion the *spectrum of a manifold*, defined in terms of eigenvalues of a generalized Laplacian. In 1964, a series of high-dimensional examples clarified the way in which the spectrum determines the object and then five years ago a breakthrough provided examples in dimension 2!

A surface sits in 3-dimensional Euclidean space and, in a neighborhood, appears to be like the (2-dimensional) plane. A manifold, M^n, is a generalized surface; that is, it looks locally like R^n, where n might be greater than 2. Any such manifold M^n can be viewed as a subset of a Euclidean space R^N of larger dimension so that it has an n-dimensional tangent space at each point in the same way that a surface in R^3 has a tangent plane at each point. The set M thus inherits an inner product on its tangent space at each point in a natural way; there are, however, many possible ways to assign an inner product to a vector space like the tangent space. Generalization: a *Riemannian manifold* is a manifold together with a "smooth" assignment of an inner product to each tangent space.

The fact that M^n is embedded in R^N makes it easy to define the concepts of vector calculus on M^n. To obtain the Laplacian Δ of a function f on M we first extend f to a function $\tilde{f}$ on an open subset of R^N and then define Δf to be the restriction to M of the ordinary Laplacian $-\sum_{i=1}^{N}\partial^2\tilde{f}/\partial x_i^2$. Geometers choose the negative sign here so that the eigenvalues of Δ are nonnegative rather than nonpositive. (In fact one way to distinguish an analyst from a geometer is to ask what the sign of Δ is.) The value of Δf will be independent of the extension $\tilde{f}$ used to define it; so it makes sense to define the *spectrum* of M to be the set of eigenvalues of Δ,

$$\{\lambda \in \mathbb{R} \mid \Delta f = \lambda f \text{ for some nonzero function } f \text{ defined on } M\}.$$

If M^n is the n sphere S^n in $\mathbb{R}^{n+1}$ for example, we may extend any f to a function $\tilde{f}$ which is constant in the radial direction and the eigenfunctions f with $\Delta f = \lambda f$ are precisely the usual spherical harmonics. In general the spectrum of M^n is a discrete set of nonnegative numbers tending to infinity, each with a finite multiplicity. For an abstractly defined Riemannian manifold M^n, the spectrum of M turns out to be independent of the way the structure is obtained by the embedding into Euclidean space.

The most primitive question which can be asked about the influence of the spectrum on the geometry of the manifold is whether or not it determines M up to isometry (rigid motion). M. Kac's beautifully written article in this MONTHLY (1966) about the Euclidean problem quotes L. Bers' interpretation of this question as "can you hear the shape of a drum?" The first result along these lines is the example John Milnor produced, in 1964, of two 16-dimensional manifolds (in fact, tori) that have the same spectrum (are "isospectral") but are not isometric. The Milnor example [**1**, p. 154] is elegant in its simplicity. Let $\{\vec{e}_1,\ldots,\vec{e}_n\}$ be a basis for R^n such that $\vec{e}_i\cdot\vec{e}_j$ is an integer for each i and j, and let Γ be the set of linear combinations of the basis using integer coefficients only. (The set Γ is called the lattice generated by the basis.) The lattice Γ is a subgroup of R^n and the quotient, $T = R^n/\Gamma$, is a manifold called a torus, which inherits the Euclidean metric. (If $\vec{e}_1 = (1,0)$ and $\vec{e}_2 = (0,1)$, this is the usual torus obtained by identifying opposite edges of a square.) If $\lambda \in \Gamma$ is fixed, then the complex-valued function on R^n defined by $x \to e^{2\pi i\gamma\cdot x}$ has the same value at x and $x+\gamma'$ (for any $\gamma' \in \Gamma$) and so gives a function on the torus R^n/Γ. As $\Delta f = -\sum\partial^2 f/\partial x_j^2$, therefore $\Delta f_\gamma = 4\pi^2|\gamma|^2 f_\gamma$, so that $\lambda = 4\pi^2|\gamma|^2$ is an eigenvalue of Δ for each γ in Γ. Note that the multiplicity of λ depends only on how many elements of Γ are located on a sphere of radius $\sqrt{\lambda}/2\pi$ about the origin in R^n. Because $\{f_\gamma \mid \gamma \in \Gamma\}$ is dense in the set of all complex-valued functions, every element of the spectrum of T must be of the form $4\pi^2|\gamma|^2$.

We see thus that the eigenvalues (and multiplicities) on a torus depend only on the number of lattice points on spheres of a given radius about the origin. On the other hand, R^n/Γ_1 and R^n/Γ_2 are isometric precisely when there is a rigid motion of R^n carrying the lattice Γ_1 onto Γ_2. However, Milnor pointed out 2 lattices discovered by E. Witt in 1941, which do not satisfy the rigid motion criterion but are such that every ball about the origin contains the same number of elements of Γ_1 as Γ_2. Using considerably more machinery (i.e., the theory of modular forms) it can be shown [**1**] that there is also a torus counterexample in dimension 12.

What then is the lowest dimension in which isospectral, nonisometric manifolds can exist? It came as quite a surprise when, in 1978, Marie-France Vignéras gave examples of 2-dimensional isospectral manifolds (actually, Riemann surfaces) which are not isometric. In fact, she gave examples in each dimension $n \geqslant 2$ of isospectral, nonisometric manifolds which don't reduce to a product of lower dimensional examples. Her 3-dimensional examples are more astounding because they have different fundamental groups. Thus spectral data do not suffice to determine even the feeble topological property of homotopy type, let alone the geometry. Her construction starts with the hyperbolic upper half plane $H = \{x + iy | y > 0\} \subset \mathbf{C}$ rather than Euclidean space as Milnor's does. A nonsingular matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with real coefficients (i.e., an element of $GL(2, R)$) acts on H as a linear fractional transformation; that is, $Az = (az + b)/(cz + d)$. A "nice" subgroup of $GL(2, R)$ will give a 2-dimensional manifold (actually, a Riemann surface) in the way that the lattice Γ produces a torus from R^n. Vignéras, by using a deep construction in quaternion algebras, is able to produce two subgroups that give isospectral manifolds which are not isometric. Details of her work are sketched in Section 2 of [**3**]. Thus we are led to the question: Does there exist a continuous, isospectral but not isometric family of Riemannian metrics? This problem seems to be quite elusive with only partial results along these lines known today. Anyone who is interested in the kinds of questions, applications, or techniques of the subject can get a good overview by a trip through [**2**]. These ideas will provide much food for thought for geometers and global analysts for many years to come.

References

1. M. Berger, P. Gauduchon, and E. Mazet, Le Spectre d'une Variété Riemannienne, Lecture Notes in Mathematics, 194, Springer-Verlag, 1971.

2. R. Osserman and A. Weinstein, Editors, Geometry of the Laplace Operator. Proc. of Symposia in Pure Mathematics, vol. 36, American Mathematical Society, 1979.

3. M-F. Vignéras, Variétés Riemanniennes Isospectrales et Non Isométriques, Ann. of Math. (2), 112 (1980) 21–32.

Added in Proof: C. Gordon and E. Wilson have announced, in abstract 803-53-017 (April, 1983) of the Notices of the AMS, a continuous isospectral change by nonisometric Riemannian metrics.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

CHARACTERIZATION OF SMOOTH DOMAINS IN C BY THEIR BIHOLOMORPHIC SELF-MAPS

STEVEN G. KRANTZ

Mathematics Department, Pennsylvania State University, University Park, PA 16802

The classification theory for Riemann surfaces contains much important and striking information about biholomorphic maps of domains $\Omega \subseteq \mathbf{C}$. For obvious reasons, material of this nature is not often presented in a first course in complex analysis. The purpose of this note is to provide a short and elementary proof of a classification of smoothly bounded plane domains which have transitive automorphism groups. The author became aware of (the general idea of) this proof in the setting of several complex variables.

What then is the lowest dimension in which isospectral, nonisometric manifolds can exist? It came as quite a surprise when, in 1978, Marie-France Vignéras gave examples of 2-dimensional isospectral manifolds (actually, Riemann surfaces) which are not isometric. In fact, she gave examples in each dimension $n \geqslant 2$ of isospectral, nonisometric manifolds which don't reduce to a product of lower dimensional examples. Her 3-dimensional examples are more astounding because they have different fundamental groups. Thus spectral data do not suffice to determine even the feeble topological property of homotopy type, let alone the geometry. Her construction starts with the hyperbolic upper half plane $H = \{x + iy | y > 0\} \subset \mathbf{C}$ rather than Euclidean space as Milnor's does. A nonsingular matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with real coefficients (i.e., an element of $GL(2, R)$) acts on H as a linear fractional transformation; that is, $Az = (az + b)/(cz + d)$. A "nice" subgroup of $GL(2, R)$ will give a 2-dimensional manifold (actually, a Riemann surface) in the way that the lattice Γ produces a torus from R^n. Vignéras, by using a deep construction in quaternion algebras, is able to produce two subgroups that give isospectral manifolds which are not isometric. Details of her work are sketched in Section 2 of [**3**]. Thus we are led to the question: Does there exist a continuous, isospectral but not isometric family of Riemannian metrics? This problem seems to be quite elusive with only partial results along these lines known today. Anyone who is interested in the kinds of questions, applications, or techniques of the subject can get a good overview by a trip through [**2**]. These ideas will provide much food for thought for geometers and global analysts for many years to come.

References

1. M. Berger, P. Gauduchon, and E. Mazet, Le Spectre d'une Variété Riemannienne, Lecture Notes in Mathematics, 194, Springer-Verlag, 1971.

2. R. Osserman and A. Weinstein, Editors, Geometry of the Laplace Operator. Proc. of Symposia in Pure Mathematics, vol. 36, American Mathematical Society, 1979.

3. M-F. Vignéras, Variétés Riemanniennes Isospectrales et Non Isométriques, Ann. of Math. (2), 112 (1980) 21–32.

Added in Proof: C. Gordon and E. Wilson have announced, in abstract 803-53-017 (April, 1983) of the Notices of the AMS, a continuous isospectral change by nonisometric Riemannian metrics.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

CHARACTERIZATION OF SMOOTH DOMAINS IN C BY THEIR BIHOLOMORPHIC SELF-MAPS

STEVEN G. KRANTZ
Mathematics Department, Pennsylvania State University, University Park, PA 16802

The classification theory for Riemann surfaces contains much important and striking information about biholomorphic maps of domains $\Omega \subseteq \mathbf{C}$. For obvious reasons, material of this nature is not often presented in a first course in complex analysis. The purpose of this note is to provide a short and elementary proof of a classification of smoothly bounded plane domains which have transitive automorphism groups. The author became aware of (the general idea of) this proof in the setting of several complex variables.

We need a little terminology. A *smooth domain* $\Omega \subseteq \mathbb{C}$ is a bounded, connected open set whose boundary consists of finitely many C^1 Jordan curves. If $\phi: \Omega \to \Omega$ is a one-to-one, onto holomorphic function (with a holomorphic inverse!) then ϕ will be called an *automorphism* of Ω.

For fixed Ω, the set of automorphisms of Ω (denoted by $\operatorname{Aut}\Omega$) forms a group under composition. Namely, the composition of two automorphisms is an automorphism, composition is associative, and the inverse of an automorphism ϕ is ϕ^{-1}. The group $\operatorname{Aut}\Omega$ is said to act *transitively* on Ω (or Ω is said to have "transitive automorphism group") if, given $p, q \in \Omega$, there is a $\phi \in \operatorname{Aut}\Omega$ such that $\phi(p) = q$. Which domains Ω have transitive automorphism group?

Every first course in complex analysis contains a discussion of the automorphism group of the unit disc D. One learns that the automorphisms of the disc consist precisely of the rotations

$$\rho_\theta(z) = e^{i\theta}z, \quad 0 \leqslant \theta < 2\pi,$$

the Möbius transformations

$$\phi_a(z) = \frac{z-a}{1-\bar{a}z}, \quad 0 \leqslant |a| < 1,$$

and compositions of these. It is easy to see that $\operatorname{Aut} D$ acts transitively on D. For if $p, q \in D$, then $\phi_{-q} \circ \phi_p$ maps p to q. The main result of this note is

THEOREM. *If $\Omega \subseteq \mathbb{C}$ is a smooth domain with transitive automorphism group, then Ω is conformally equivalent to the disc.*

The proof will take up the rest of this note and will be broken up into several lemmas. The proof reveals a nice interplay between function-algebraic ideas and geometric ones.

DEFINITION. A function $f: \bar{\Omega} \to \mathbb{C}$ will be called a *peaking function* for the point $P \in \partial\Omega$ if f is continuous on $\bar{\Omega}$, f is holomorphic on Ω, $f(P) = 1$, and $|f(z)| < 1$ for $z \in \bar{\Omega}\setminus\{P\}$. The point P is then called a *peak point*.

LEMMA 1. *Given a smooth domain Ω, there is a peak point $P \in \partial\Omega$ and a peaking function f for the point P.*

Proof. Let $P \in \partial\Omega$ be a point in $\bar{\Omega}$ which is furthest from 0. Then

$$f(z) = \frac{1}{2}\left(\frac{z\bar{P}}{|P|^2} + 1\right)$$

does the job. □

LEMMA 2. *Let $P \in \partial\Omega$ be a peak point. Let $z_0 \in \Omega$ be fixed. Suppose that $\phi_j \in \operatorname{Aut}\Omega$ satisfy $\phi_j(z_0) \to P$ as $j \to \infty$. If $K \subset \Omega$ is a compact set, then $\phi_j(z) \to P$ uniformly for $z \in K$.*

Proof. Let f be a peak function for Ω at P. Consider the family $\{f \circ \phi_j\}$. Since this family of functions is bounded by 1, it is normal. By Montel's Theorem, every subsequence has a subsequence which converges uniformly on compact sets to a holomorphic function. But $(f \circ \phi_j)(z_0) \to 1$. By the Maximum Modulus Principle, the only possible limit function is constant, so $f \circ \phi_j$ converges uniformly on K to the constant function 1. Since f takes the value 1 only at P, the proof is complete. □

LEMMA 3. *If Ω has transitive automorphism group, then Ω is simply connected.*

Proof. If Ω is not simply connected, then there is a loop $\lambda: [0,1] \to \Omega$ which cannot be continuously deformed to a point in Ω. Let $P \in \partial\Omega$ be a peak point. Let $z_j \in \Omega$ satisfy $z_j \to P$. Fix $z_0 \in \Omega$. Since $\operatorname{Aut}\Omega$ is transitive, there exist $\phi_j \in \operatorname{Aut}\Omega$ such that $\phi_j(z_0) = z_j$. Since $\Lambda \equiv \{\lambda(t): 0 \leqslant t \leqslant 1\}$ is compact, Lemma 2 implies that $\phi_j(z) \to P$ uniformly for $z \in \Lambda$.

Now choose a neighborhood U of P such that $U \cap \Omega$ is simply connected (since $\partial\Omega$ is nice, we

can do this—see Fig. 1). By the definition of uniform convergence, there is a j so large that $\phi_j(\Lambda) \subseteq U \cap \Omega$. Since $U \cap \Omega$ is simply connected, $\phi_j(\Lambda)$ can be continuously deformed to a point in $U \cap \Omega$. But since ϕ_j is a homeomorphism, Λ can be continuously deformed to a point. That is a contradiction. □

Now the proof of the theorem is immediate. For Lemma 3 says, together with the Riemann

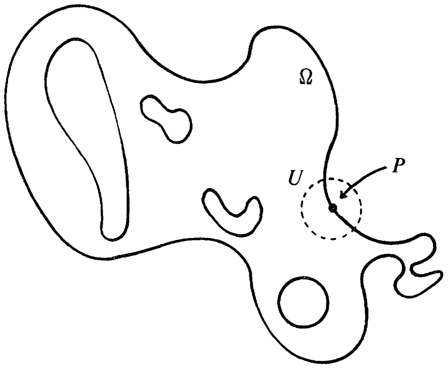


FIG. 1

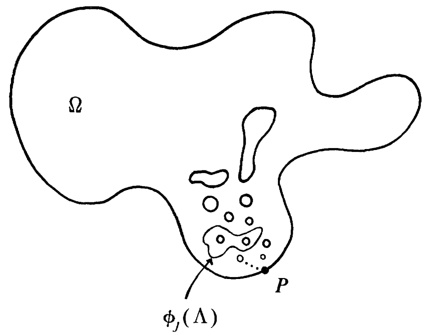


FIG. 2

Mapping Theorem, that a smooth domain with transitive automorphism group is conformally equivalent to the disc.

It is not necessary to assume that $\partial\Omega$ is nice in order to prove this result. But the proof we have given will not work in general. For if Ω has infinitely many holes which accumulate at P (see Fig. 2), then there may be no neighborhood U of P such that $U \cap \Omega$ is simply connected. So $\phi_j(\Lambda)$ may not be deformable to a point, no matter how large j.

Our theorem is an old one which is known to experts in one complex variable. It has been extended to the setting of several complex variables by Bun Wong and Rosay. It is discussed in Section 10.2 of [**1**].

Reference

1. S. G. Krantz, Function Theory of Several Complex Variables, Wiley, 1982.

THE SEQUENCE OF DERIVATIVES OF A C^∞ FUNCTION

MICHAEL J. HOFFMAN AND RICHARD KATZ

Department of Mathematics and Computer Science, California State University, Los Angeles, CA 90032

It is an interesting exercise in a course on distributions and generalized functions to show that the assignment of $\Sigma f^{(n)}(0)$ to f is not a distribution on the space of infinitely differentiable functions of compact support on the real line. A solution is to multiply something like the exponential function by a "bump" function of compact support which is constantly equal to 1 on a neighborhood of 0 to create an f for which the series does not converge. One can actually prove much more without using such a "bump" function. If f is any nonzero, infinitely differentiable, compactly supported function, there must be a point x where the sequence $f^{(n)}(x)$ does not converge to 0. In fact there must be sequences of points converging to each endpoint of the support of f at each point of which the derivatives of f fail to remain bounded. This assertion will be derived from the following proposition about the relationship between analyticity and infinite differentiability. The proof of the assertion itself is a simple illustration of the use of complex methods to obtain real results. We conclude with some remarks about the situation for functions

can do this—see Fig. 1). By the definition of uniform convergence, there is a j so large that $\phi_j(\Lambda) \subseteq U \cap \Omega$. Since $U \cap \Omega$ is simply connected, $\phi_j(\Lambda)$ can be continuously deformed to a point in $U \cap \Omega$. But since ϕ_j is a homeomorphism, Λ can be continuously deformed to a point. That is a contradiction. □

Now the proof of the theorem is immediate. For Lemma 3 says, together with the Riemann

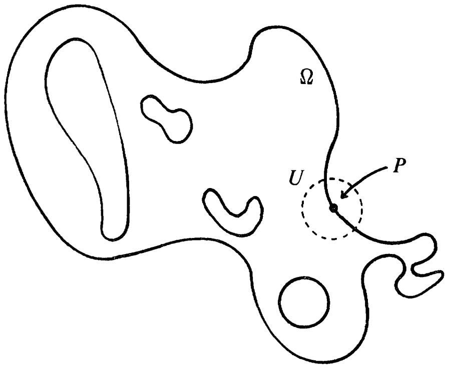


FIG. 1

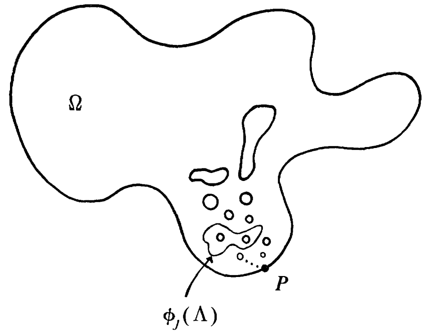


FIG. 2

Mapping Theorem, that a smooth domain with transitive automorphism group is conformally equivalent to the disc.

It is not necessary to assume that $\partial\Omega$ is nice in order to prove this result. But the proof we have given will not work in general. For if Ω has infinitely many holes which accumulate at P (see Fig. 2), then there may be no neighborhood U of P such that $U \cap \Omega$ is simply connected. So $\phi_j(\Lambda)$ may not be deformable to a point, no matter how large j.

Our theorem is an old one which is known to experts in one complex variable. It has been extended to the setting of several complex variables by Bun Wong and Rosay. It is discussed in Section 10.2 of [**1**].

Reference

1. S. G. Krantz, Function Theory of Several Complex Variables, Wiley, 1982.

THE SEQUENCE OF DERIVATIVES OF A C^∞ FUNCTION

MICHAEL J. HOFFMAN AND RICHARD KATZ
Department of Mathematics and Computer Science, California State University, Los Angeles, CA 90032

It is an interesting exercise in a course on distributions and generalized functions to show that the assignment of $\Sigma f^{(n)}(0)$ to f is not a distribution on the space of infinitely differentiable functions of compact support on the real line. A solution is to multiply something like the exponential function by a "bump" function of compact support which is constantly equal to 1 on a neighborhood of 0 to create an f for which the series does not converge. One can actually prove much more without using such a "bump" function. If f is any nonzero, infinitely differentiable, compactly supported function, there must be a point x where the sequence $f^{(n)}(x)$ does not converge to 0. In fact there must be sequences of points converging to each endpoint of the support of f at each point of which the derivatives of f fail to remain bounded. This assertion will be derived from the following proposition about the relationship between analyticity and infinite differentiability. The proof of the assertion itself is a simple illustration of the use of complex methods to obtain real results. We conclude with some remarks about the situation for functions

in the Schwartz space of C^∞ functions of rapid decrease.

If f is infinitely differentiable at a point t, let $r(f;\, t)$ be the radius of convergence of the formal Taylor series $\sum_{n=0}^{\infty}(f^{(n)}(t)/n!)(x-t)^n$. Recall that f is analytic at t if this series converges to $f(x)$ on some neighborhood of t. Let $C^\infty(a, b)$ be the functions infinitely differentiable on the open interval (a, b) and let $C_0^\infty(a, b)$ be the infinitely differentiable functions defined on the real line with compact support contained in the closed interval $[a, b]$.

PROPOSITION. *Suppose $f \in C^\infty(a, b)$ and $r(f;\, t) > \delta > 0$ for all $t \in (a, b)$. Then f is analytic on (a, b).*

This proposition may be found in [**1**, p. 183] and is proved in [**4**, Hilfs. 6]. We give here a proof which may be accessible to upper division mathematics students and shows the use of the Baire category theorem.

LEMMA 1. *Suppose $f \in C^\infty(c, d)$ and that for some N,*

$$\sup\{|f^{(n)}(s)/n!|^{1/n} : c < s < d \text{ and } n \geqslant N\} < \infty.$$

Then f is analytic on the open interval (c, d).

Proof. Let M be the indicated supremum. If x and t are in (c, d) and $|x - t| < 1/M$, Taylor's formula with remainder assures us that

$$f(x) = \sum_{k=0}^{n} (f^{(k)}(t)/k!)(x-t)^k + (f^{(n+1)}(s)/(n+1)!)(x-t)^{n+1}$$

for some point s between x and t. If $n > N$, the remainder term is bounded by $|(x-t)M|^{n+1}$. Since this goes to 0 as n grows, we have the desired convergence and f is analytic at t. ∎

LEMMA 2. *Suppose $f \in C^\infty(a, b)$; that $a < c < d < b$; and that f is analytic on (c, d). Suppose $r(f;\, t) > \delta$ for all t in $[c, d]$ and that $c < x < d$. If $|x - c| < \delta$, then*

$$f(x) = \sum (f^{(k)}(c)/k!)(x-c)^k.$$

If $|x - d| < \delta$, then

$$f(x) = \sum (f^{(k)}(d)/k!)(x-d)^k.$$

Proof. We prove the first case. The second is similar. Pick t between c and x. The series

$$g(y) = \sum (f^{(n)}(t)/n!)(y-t)^n$$

defines a function g analytic for $t - \delta < y < t + \delta$. Since f is analytic on (c, d), g and f and all their derivatives agree on (c, x). Since they are continuous, they also agree at c. Thus

$$f(x) = g(x) = \sum (g^{(n)}(c)/n!)(x-c)^n = \sum (f^{(n)}(c)/n!)(x-c)^n. \blacksquare$$

Proof of the Proposition. Suppose f is not analytic at a point x_0 in (a, b). Choose $[a', b'] \subset (a, b)$ with x_0 in (a', b'), and let $F = \{x \in [a', b']: f \text{ is not analytic at } x\}$. Then F is a nonempty complete metric space. By the Cauchy-Hadamard formula for $r(f;\, x)$, the sets

$$E_N = \{x \in [a', b'] : |f^{(n)}(x)/n!|^{1/n} \leqslant 1/\delta \text{ for all } n \geqslant N\}$$

exhaust $[a'b']$. By the Baire category theorem, one of the closed sets E_N has a nonempty interior relative to F. So there is an open interval I such that $\varnothing \neq I \cap F \subseteq E_N$. Without loss of generality, we can assume that I is contained in (a, b) and has length less than $\delta/2$. Any point s in $I \setminus F$ is contained in an open interval (c, d) on which f is analytic and one of whose endpoints, say c, is in $I \cap F$. By Lemma 2,

$$f(s)=\sum_{n=0}^{\infty}(f^{(n)}(c)/n!)(s-c)^n.$$

So, for $k>N$:

$$\begin{aligned}|f^{(k)}(s)| &= \left|\sum_{j=k}^{\infty}(f^{(j)}(c)/j!)j(j-1)\cdots(j-k+1)(s-c)^{j-k}\right| \\ &\leqslant \sum_{j=k}^{\infty}(1/\delta)^j j(j-1)\cdots(j-k+1)(\delta/2)^{j-k} \\ &= (2/\delta)^k\sum_{j=k}^{\infty}j(j-1)\cdots(j-k+1)(1/2)^j \\ &= (2/\delta)^k(d/dx)^k(1/(1-x))|_{1/2} \\ &= (2/\delta)^k(1/(1-(1/2)))^{k+1}k!\end{aligned}$$

Thus $|f^{(k)}(s)/k!|^{1/k}$ is bounded by $1/\delta$ on $F\cap I$ for $k>N$, and by $8/\delta$ on the rest of I. Lemma 1 shows that f is analytic on I contrary to the assumption that there were points of F in I. ∎

We now derive our original assertion from this proposition. Suppose f is a nonzero function in $C_0^\infty(a,b)$ and that a is the infimum of the support of f. If the derivatives $f^{(n)}(t)$ remain bounded at every point in some interval (a,c), then $r(f;t)=+\infty$ for all t in $(-\infty,c)$. The proposition shows that f extends from $(-\infty,c)$ to an entire function which is identically 0 on $(-\infty,a)$. The identity theorem for analytic functions shows that f would be 0 everywhere which is impossible because $a=\inf\operatorname{supp}(f)$. This contradiction gives the desired result.

The situation is quite different if we relax the condition that f have compact support and require only that it be in the Schwartz space, $\mathbb{S}$, of functions of rapid decrease. Thus f and all its derivatives should go to 0 as $|x|\to\infty$ even when multiplied by any polynomial. There are functions in $\mathbb{S}$ for which the sequence of derivatives converges uniformly to 0. The key fact needed is that the Fourier transform and its inverse take $\mathbb{S}$ onto $\mathbb{S}$ [**3**, Th. IX.1]. Choose F in $C_0^\infty(-1,1)$ to be an even function with nonnegative real values. Put

$$f(z)=\check{F}(z)=(2\pi)^{-1/2}\int_{-1}^{1}F(t)e^{izt}dt.$$

Then f is in $\mathbb{S}$, and for real x and $k\geqslant 0$, we have

$$f^{(k)}(x)=(2\pi)^{-1/2}\int_{-1}^{1}F(t)e^{ixt}(it)^k dt.$$

These are real, and $|f^{(k)}(x)|\leqslant C/(k+1)$ for a constant C which depends on f but not on x. So, for such f, the derivatives converge uniformly to 0.

There are restrictions on such a function. If $f\in C^\infty(\mathbb{R})$ and the sequence of derivatives is bounded at every point, the proposition shows that f extends to an entire function. Examination of the Taylor series around 0 shows that if $|f^{(k)}(0)|\leqslant M$ for all k, then $|f(z)|\leqslant Me^{|z|}$. So f is the restriction of an entire function of exponential type 1. This means that the function cannot decay too rapidly along the real axis and it must grow fairly rapidly in the complex plane if it is to be in $\mathbb{S}$. In particular it cannot grow more slowly than $e^{|z|^{1/2}}$ since an entire function of order less than 1/2 cannot be bounded along any ray [**2**, Th. 3.1.5]. Since the function is of order at most 1, it cannot go to 0 faster than exponentially along the real axis [**2**, Cor. 5.1.14]. It can go to 0 more slowly and still be in $\mathbb{S}$. If it is, then it is certainly in $L^2(\mathbb{R})$ and the Paley-Wiener theorem says that f is the inverse Fourier transform of a C^∞ function supported in $[-1,1]$. So the construction above gives all such functions in $\mathbb{S}$.

In summary we have: If $f\in C_0^\infty(\mathbb{R})$, then there are sequences of points converging to each

endpoint of the support of f at each point of which the sequence of derivatives of f is unbounded. This is not true for all f in $\mathbb{S}$. However, for $f \in \mathbb{S}$, the sequence of derivatives is bounded at every point if and only if f is the inverse Fourier transform of a function in $C_0^\infty(-1, 1)$ and, in that case, the sequence of derivatives converges uniformly to 0. In fact, $|f^{(k)}(x)| = O(1/(k+1))$ uniformly in x.

References

1. R. P. Boas, Jr., A Primer of Real Functions, 3rd ed., The Carus Mathematical Monographs, No. 13, The Mathematical Association of America, Washington, DC, 1981.
2. R. P. Boas, Jr., Entire Functions, Academic Press, New York, 1954.
3. M. Reed and B. Simon, Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness, Academic Press, New York, 1975.
4. H. Salzmann and K. Zeller, Singularitäten unendlich oft differenzierbarer Funktionen, Math. Z., 62 (1955) 354–367.

AN INEQUALITY FOR VARIATIONS

P. S. Bullen
Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

Recently Russell [1] has given a simple proof of the fact that under certain boundary conditions the total variation of the product of two functions of bounded variation is less than or equal to the product of their total variations. More precisely, if f and g belong to $BV^*[a, b]$, where $BV^*[a, b] = \{h;\ h \text{ is of bounded variation on } [a, b] \text{ and } h(a) = 0\}$, then $V(fg) \leqslant V(f)V(g)$, where $V(f)$ denotes the total variation of f on $[a, b]$.

The following is an even simpler proof of this theorem that depends on the following elementary facts.

(a) If $f \in BV^*[a, b]$, then f can be decomposed as $f = f_1 - f_2$, where f_1 and f_2 are increasing and $f_1(a) = f_2(a) = 0$.

(b) A particular example of such a decomposition is when f_1 and f_2 are the positive and negative variations of f, respectively; let us call this the canonical decomposition of f.

(c) For all such decompositions of f, $f_1(b) + f_2(b) \geqslant V(f)$, with equality in the case of the canonical decomposition.

Our proof then runs as follows. Let $f = f_1 - f_2$, $g = g_1 - g_2$, be the canonical decompositions of f and g respectively; then

$$fg = (f_1 g_1 + f_2 g_2) - (f_1 g_2 + f_2 g_1),$$

and the right-hand side is a decomposition of fg.

Hence, by (c)

$$\begin{aligned} V(fg) &\leqslant (f_1(b)g_1(b) + f_2(b)g_2(b)) + (f_1(b)g_2(b) + f_2(b)g_1(b)) \\ &= (f_1(b) + f_2(b))(g_1(b) + g_2(b)) \\ &= (f_1 + f_2)(g_1 + g_2) \\ &= V(f)V(g). \end{aligned}$$

Reference

1. A. M. Russell, A commutative Banach algebra of functions of bounded variation, this Monthly, 87 (1980) 39–40.

endpoint of the support of f at each point of which the sequence of derivatives of f is unbounded. This is not true for all f in $\mathbb{S}$. However, for $f \in \mathbb{S}$, the sequence of derivatives is bounded at every point if and only if f is the inverse Fourier transform of a function in $C_0^\infty(-1, 1)$ and, in that case, the sequence of derivatives converges uniformly to 0. In fact, $|f^{(k)}(x)| = O(1/(k+1))$ uniformly in x.

References

1. R. P. Boas, Jr., A Primer of Real Functions, 3rd ed., The Carus Mathematical Monographs, No. 13, The Mathematical Association of America, Washington, DC, 1981.

2. R. P. Boas, Jr., Entire Functions, Academic Press, New York, 1954.

3. M. Reed and B. Simon, Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness, Academic Press, New York, 1975.

4. H. Salzmann and K. Zeller, Singularitäten unendlich oft differenzierbarer Funktionen, Math. Z., 62 (1955) 354–367.

AN INEQUALITY FOR VARIATIONS

P. S. Bullen
Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

Recently Russell [1] has given a simple proof of the fact that under certain boundary conditions the total variation of the product of two functions of bounded variation is less than or equal to the product of their total variations. More precisely, if f and g belong to $BV^*[a, b]$, where $BV^*[a, b] = \{h;\ h \text{ is of bounded variation on } [a, b] \text{ and } h(a) = 0\}$, then $V(fg) \leqslant V(f)V(g)$, where $V(f)$ denotes the total variation of f on $[a, b]$.

The following is an even simpler proof of this theorem that depends on the following elementary facts.

(a) If $f \in BV^*[a, b]$, then f can be decomposed as $f = f_1 - f_2$, where f_1 and f_2 are increasing and $f_1(a) = f_2(a) = 0$.

(b) A particular example of such a decomposition is when f_1 and f_2 are the positive and negative variations of f, respectively; let us call this the canonical decomposition of f.

(c) For all such decompositions of f, $f_1(b) + f_2(b) \geqslant V(f)$, with equality in the case of the canonical decomposition.

Our proof then runs as follows. Let $f = f_1 - f_2$, $g = g_1 - g_2$, be the canonical decompositions of f and g respectively; then

$$fg = (f_1g_1 + f_2g_2) - (f_1g_2 + f_2g_1),$$

and the right-hand side is a decomposition of fg.

Hence, by (c)

$$\begin{aligned} V(fg) &\leqslant (f_1(b)g_1(b) + f_2(b)g_2(b)) + (f_1(b)g_2(b) + f_2(b)g_1(b)) \\ &= (f_1(b) + f_2(b))(g_1(b) + g_2(b)) \\ &= (f_1 + f_2)(g_1 + g_2) \\ &= V(f)V(g). \end{aligned}$$

Reference

1. A. M. Russell, A commutative Banach algebra of functions of bounded variation, this Monthly, 87 (1980) 39–40.

References

1. L. Dubikajtis, C. Ferens, R. Ger, and M. Kuczma, On Mikusinski's functional equation, Ann. Polon. Math., 28 (1973) 39–47.
2. M. E. Kuczma and M. Kuczma, An elementary proof and an extension of a theorem of Steinhaus, Glasnik Mat., 6 (26) (1971) 11–18.
3. S. Piccard, Sur des Ensembles Parfaits, Paris, 1942.
4. H. Steinhaus, Sur les distances des points des ensembles de mesure positive, Fund. Math., 1 (1920) 99–104.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

SERENDIPITY IN MATHEMATICS OR HOW ONE IS LED TO DISCOVER THAT

$$\sum_{n=1}^{\infty} \frac{1\cdot 3\cdot 5\cdots(2n-1)}{n2^n n!} = \frac{1}{2} + \frac{3}{16} + \frac{15}{144} + \cdots = \ln 4$$

BERTRAM ROSS

Department of Mathematics, University of New Haven, West Haven, CT 06516

This paper chronicles the course of a mathematical discovery. It is usually the case in (alleged) accidental discovery that the starting point is far afield from the particular discovery. Our starting point is the evaluation of the integral

$$\frac{1}{\Gamma(\nu)}\int_0^x (x-t)^{\nu-1}\ln t\,dt, \quad \mathrm{Re}(\nu) \geqslant 0. \tag{1}$$

An integral of the form $\int_c^x (x-t)^{\nu-1} f(t)\,dt/\Gamma(\nu)$ is known as the Riemann-Liouville integral which is the cornerstone of the *fractional calculus*. More precisely stated, it is the integral which defines integration and differentiation to an arbitrary order. The R-L integral can be denoted by ${}_cD_x^{-\nu}f(x)$, a notation devised by Harold T. Davis [**1**]. The subscripts on D denote the terminals of integration. The order of differintegration [**2**] can be rational, irrational or complex. When $\nu = n$, an integer, then ${}_0D_x^{-\nu}f(x)$ and ${}_0D_x^{\nu}f(x)$ are respectively ordinary integration and differentiation.

The idea of differentiation to an arbitrary order started in 1695 when L'Hôpital asked Leibniz what would happen with $d^n y/dx^n$ if $n = 1/2$. Hence, the topic started with the misnomer fractional calculus [**3**].

One method of evaluating (1) is a series expansion followed by term by term integration. Let

$$t = x + t - x = x\left(1 + \frac{t-x}{x}\right), \tag{2}$$

where x and t are real and $x > 0$. Then

$$\ln t = \ln x + \ln\left(1 + \frac{t-x}{x}\right).$$

When $|(t-x)/x| < 1$, we can expand $\ln(1 + (t-x)/x)$ in a Taylor's series expansion:

$$\ln t = \ln x + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}(t-x)^n}{nx^n} = \ln x - \sum_{n=1}^{\infty} \frac{(x-t)^n}{nx^n}, \tag{3}$$

where the interval of convergence is $0 < t \leqslant 2x$.

When (3) is substituted into (1) we have

$$ {}_0D_x^{-\nu}\ln x = \frac{\ln x}{\Gamma(\nu)}\int_0^x (x-t)^{\nu-1}\,dt - \frac{1}{\Gamma(\nu)}\int_0^x (x-t)^{\nu-1}\sum_{n=1}^{\infty}\frac{(x-t)^n}{nx^n}\,dt. $$

Thus, for integration of ln x to an arbitrary order we get

$$ \text{(4)}\qquad {}_0D_x^{-\nu}\ln x = \frac{x^{\nu}\ln x}{\nu\Gamma(\nu)} - \frac{x^{\nu}}{\Gamma(\nu)}\sum_{n=1}^{\infty}\frac{1}{n(\nu+n)},\quad \operatorname{Re}(\nu) \geqslant 0. $$

Because of the property of analyticity, we can interchange $-\nu$ and ν. So, for differentiation of ln x to an arbitrary order we will have

$$ \text{(5)}\qquad {}_0D_x^{\nu}\ln x = \frac{x^{-\nu}\ln x}{-\nu\Gamma(-\nu)} - \frac{x^{-\nu}}{\Gamma(-\nu)}\sum_{n=1}^{\infty}\frac{1}{n(-\nu+n)},\quad \operatorname{Re}(\nu) \geqslant 0. $$

The results in (4) and (5) can be shown to be consistent with the results of ordinary integration and differentiation for the case $\nu = 1$.

This is where the matter stood for some time until there was sufficient time and inclination to pursue this work further. It seemed like a good idea to write both terms on the right sides of (4) and (5) with the same denominator. This requires the multiplication of the second term on the right in (4) by unity in the form of ν/ν and in (5) by $-\nu/-\nu$. It was after playing with symbols and naively placing ν and $-\nu$ under the summation signs that pattern recognition occurred and the adrenalin started to flow.

From the study of the psi function [4] we learn that series of the form

$$ \sum_{n=1}^{\infty}\frac{y}{n(y+n)} = C + \Psi(1+y), $$

where C is Euler's constant and Ψ is the psi function, tables of which can be found in [5]. In terms of the psi function (4) and (5) become respectively

$$ \text{(6)}\qquad {}_0D_x^{-\nu}\ln x = \frac{x^{\nu}}{\Gamma(1+\nu)}[\ln x - C - \Psi(1+\nu)],\quad \operatorname{Re}(\nu) \geqslant 0, $$

and

$$ \text{(7)}\qquad {}_0D_x^{\nu}\ln x = \frac{x^{-\nu}}{\Gamma(1-\nu)}[\ln x - C - \Psi(1-\nu)],\quad \operatorname{Re}(\nu) \geqslant 0. $$

So far, nothing very spectacular has developed. The second step on the path toward mathematical discovery is the solution to an integral equation of the Volterra type. Armed with a formula for the derivative of arbitrary order of ln x, I proceeded to solve

$$ \text{(8)}\qquad \frac{1}{\Gamma(1/2)}\int_0^x (x-t)^{-1/2}f(t)\,dt = \ln x, $$

where $\nu = 1/2$ was chosen arbitrarily.

The integral is the Riemann-Liouville integral and the integral equation can be solved by Laplace transforms. However, the use of the operator notation of the fractional calculus shows its elegance and power by virtue of its simplicity and succinctness. An interested reader can compare the following with the method of Laplace transforms. We write (8) as

$$ \text{(9)}\qquad {}_0D_x^{-1/2}f(x) = \ln x. $$

Operating on both sides of the above with ${}_0D_x^{1/2}$ gives

$$ \text{(10)}\qquad f(x) = {}_0D_x^{1/2}\ln x. $$

The use of formula (7) with $\nu = 1/2$ gives us at once

$$f(x) = \frac{x^{-1/2}}{\Gamma(1/2)}[\ln x - C - \Psi(1/2)]. \tag{11}$$

Also, from the list of special values of the psi function [6] we see that

$$\Psi(1/2) = -C - \ln 4. \tag{12}$$

The solution to (8) is then

$$f(x) = \frac{x^{-1/2}}{\sqrt{\pi}} \ln 4x. \tag{13}$$

The third step on our path is where the "serendipitous" event occurs and which is given in the title. Many hours were wasted because of slight mistakes in sign and algebra, not to speak of a ream of paper in the waste basket. I could not let (13) rest without verifying it.

The solution (13) is substituted into (8) and we have to show

$$\frac{1}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2} \ln 4t \, dt = \ln x,$$

that is,

$$\frac{\ln 4}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2}\, dt + \frac{1}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2} \ln t \, dt = \ln x. \tag{14}$$

The expansion for $\ln t$ given in (3) is substituted into (14) and we have

$$\begin{aligned} &\frac{\ln 4}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2}\, dt + \frac{\ln x}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2}\, dt \\ &\quad - \frac{1}{\pi}\int_0^x (x-t)^{-1/2} t^{-1/2} \sum_{n=1}^{\infty} \frac{(x-t)^n}{n x^n}\, dt = \ln x. \end{aligned} \tag{15}$$

We recall the beta integral

$$\int_0^x (x-t)^d t^a \, dt = \frac{\Gamma(d+1)\Gamma(a+1)}{\Gamma(d+a+2)} x^{d+a+1}, d > -1, a > -1. \tag{16}$$

The first and second integrals in (15) have the values $\ln 4$ and $\ln x$ respectively. The third integral is integrated term by term. The use of formula (16) and the recurrence formula for the gamma function yields

$$\left[\frac{1}{2\cdot 1!} + \frac{1\cdot 3}{2\cdot 2^2\cdot 2!} + \frac{1\cdot 3\cdot 5}{3\cdot 2^3\cdot 3!} + \cdots\right].$$

In summation form the value of the third integral in (15) is

$$\sum_{n=1}^{\infty} \frac{1\cdot 3\cdot 5\cdots(2n-1)}{n 2^n n!}. \tag{17}$$

In order to satisfy (15) we "discover" that

$$\sum_{n=1}^{\infty} \frac{1\cdot 3\cdot 5\cdots(2n-1)}{n 2^n n!} = \ln 4. \tag{18}$$

The matter does not end at this point. The series will be put into a computer. Before doing so, and to guard against error in the development of (18), the series is tested for convergence. When the series is put into the form whose general term is $(1/\sqrt{\pi})\Gamma(n+1/2)/n\Gamma(n+1)$, d'Alembert's ratio test fails but Raabe's test is decisive. The series is then put into the computer without trying

to improve on the rapidity of convergence and without regard to the rate of error propagation. After 682,783 terms, at which point the difference between consecutive terms is less than 10^{-12}, the series summed to $1.385\ldots$, whereas the known value for $\ln 4$ is $1.386\ldots$. [8]

The relation (18) was obtained by deduction. So, as one thing leads to another, the thought arose that once a relation is discovered indirectly it might be a good idea to prove it directly. A starting point is the expression

$$\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n}$$

which is almost like the identity

$$\Gamma(n+1/2)=\frac{1\cdot 3\cdot 5\cdots(2n-1)\sqrt{\pi}}{2^n}. \text{ [7]}$$

The term $\Gamma(n+1/2)$ also appears in the evaluation of a definite integral. From the study of the beta integral we have

$$\int_0^{\pi/2}\sin^{2n}x\,dx=\frac{1}{2}\frac{\Gamma(n+1/2)\Gamma(1/2)}{\Gamma(n+1)}=\frac{\sqrt{\pi}}{2}\frac{\Gamma(n+1/2)}{n!}.$$

Denoting the series (17) by S we can write

$$S=\frac{2}{\pi}\sum_{n=1}^{\infty}\int_0^{\pi/2}\frac{\sin^{2n}x}{n}dx.$$

We can, in this instance, interchange the summation sign and the integral sign getting

$$S=\frac{2}{\pi}\int_0^{\pi/2}\left(\frac{\sin^2x}{1}+\frac{\sin^4x}{2}+\frac{\sin^6x}{3}+\cdots\right)dx.$$

Observing that the integrand above is the Maclaurin series expansion for $-\ln(1-\sin^2x)$, it follows that

$$\begin{aligned}S&=-\frac{4}{\pi}\int_0^{\pi/2}\ln\cos x\,dx\\&=-\frac{4}{\pi}\left(-\frac{\pi}{2}\ln 2\right), \quad \text{[9]}\\&=\ln 4.\end{aligned}$$

There are several other series expansions for $\ln 4$ that can be found in the literature. The result (18) is not very important in the scheme of things although the result might be of use in accelerating the rapidity of convergence of a similar series because its sum is known. The purpose of this paper is not to introduce a new result. The primary purposes are to present an exposition of techniques that could be more widely known and to exemplify the process of mathematical discovery.

The result (18) was unexpected but could not have been achieved without ability and a prepared mind. In mathematics the word serendipity does not have the ordinary meaning "lucky accident."

References

1. Harold T. Davis, The Theory of Linear Operators, Principia, Bloomington, IN, 1936.
2. Keith B. Oldham and J. Spanier, The Fractional Calculus, Academic Press, NY, 1974.
3. Bertram Ross, A Brief History and Exposition of the Fractional Calculus and Its Applications, pp. 1–36, Springer-Verlag, Lecture Notes in Mathematics, #457, Berlin, 1975.
4. ______, The Psi function, Mathematics Magazine, vol. 51, no. 3, May 1978, pp. 176–179.
5. Handbook of Mathematical Functions, National Bureau of Standards, Applied Math. Series #55, Washington, DC, 1972, p. 258.

6. Harold T. Davis, The Summation of Series, The Principia Press, San Antonio, 1962, p. 34.
7. Bertram Ross, Solved Problems in Analysis, Dover Publications, NY, 1971, p. 14.
8. Programmed by Dr. Michael J. Fischer, Yale University, on an Apollo computer.
9. Murray Spiegel, Mathematical Handbook, Schaum's Series, McGraw-Hill, NY, 1968, pp. 169–170.
10. ______, Advanced Calculus, Schaum's Series, McGraw-Hill, NY, 1963, p. 275.

PROBLEMS AND SOLUTIONS

Send all ***proposed*** *problems, typed and in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by February 29, 1984. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3013. *Proposed by Stanley Rabinowitz, Digital Equipment Corp.*

Let ABC be a fixed triangle in the plane. Let T be the transformation of the plane that maps a point P into its isotomic conjugate (relative to ABC). Let G be the transformation that maps P into its isogonal conjugate. Prove that the mappings TG and GT are affine collineations (linear transformations).

E 3014. *Proposed by Lorraine L. Foster, California State University, Northridge.*

Prove that the congruence $3^x \equiv 19 \pmod{2^n}$ is solvable for $n \geqslant 1$.

E 3015. *Proposed by Ioan Tomescu, University of Bucharest.*

Let $h \geqslant 2$ be an integer and let $\alpha(h) \geqslant h$ be the unique solution of the equation

$$1 + \frac{x}{x-1} + \frac{x}{x-2} + \cdots + \frac{x}{x-h+1} = \ln\frac{(x)_h}{h!}$$

where $(x)_h = x(x-1)\cdots(x-h+1)$. Show that

$$\lim_{h\to\infty}\alpha(h)\left(2h + \frac{1}{2}\ln h\right)^{-1} = 1.$$

6. Harold T. Davis, The Summation of Series, The Principia Press, San Antonio, 1962, p. 34.
7. Bertram Ross, Solved Problems in Analysis, Dover Publications, NY, 1971, p. 14.
8. Programmed by Dr. Michael J. Fischer, Yale University, on an Apollo computer.
9. Murray Spiegel, Mathematical Handbook, Schaum's Series, McGraw-Hill, NY, 1968, pp. 169–170.
10. ______, Advanced Calculus, Schaum's Series, McGraw-Hill, NY, 1963, p. 275.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, typed and in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by February 29, 1984. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).

E 3013. *Proposed by Stanley Rabinowitz, Digital Equipment Corp.*

Let ABC be a fixed triangle in the plane. Let T be the transformation of the plane that maps a point P into its isotomic conjugate (relative to ABC). Let G be the transformation that maps P into its isogonal conjugate. Prove that the mappings TG and GT are affine collineations (linear transformations).

E 3014. *Proposed by Lorraine L. Foster, California State University, Northridge.*

Prove that the congruence $3^x \equiv 19 \pmod{2^n}$ is solvable for $n \geqslant 1$.

E 3015. *Proposed by Ioan Tomescu, University of Bucharest.*

Let $h \geqslant 2$ be an integer and let $\alpha(h) \geqslant h$ be the unique solution of the equation

$$1+\frac{x}{x-1}+\frac{x}{x-2}+\cdots+\frac{x}{x-h+1}=\ln\frac{(x)_h}{h!}$$

where $(x)_h = x(x-1)\cdots(x-h+1)$. Show that

$$\lim_{h\to\infty}\alpha(h)\left(2h+\frac{1}{2}\ln h\right)^{-1}=1.$$

E 3016. *Proposed by Eugene Levine, Adelphi University.*

Let $a_1, a_2, \ldots, a_n$ be positive numbers ($n > 2$), let $s = a_1 + a_2 + \cdots + a_n$, and let $0 < \beta \leqslant 1$. Prove that

$$\sum_{k=1}^{n}\left(\frac{s-a_k}{a_k}\right)^{\beta} \geqslant (n-1)^{2\beta}\sum_{k=1}^{n}\left(\frac{a_k}{s-a_k}\right)^{\beta}.$$

Further, show that equality holds iff $a_1 = a_2 = \cdots = a_n$.

E 3017. *Proposed by Charles W. Schelin, University of Wisconsin, La Crosse.*

Let $f(z)$ be a real polynomial of degree $n \geqslant 1$ such that $f(-1)f(1) \neq 0$. Put $L = -f'(-1)/f(-1)$ and $R = f'(1)/f(1)$. Show that:

(a) If $L \leqslant n/2$ or $R \leqslant n/2$, then f has a zero in $|z| \geqslant 1$.
(b) If $L \geqslant n/2$ or $R \geqslant n/2$, then f has a zero in $|z| \leqslant 1$.

E 3018. *Proposed by A. M. Nadel, Pomona College.*

A nonempty set S of positive integers is said to be eventually linear if there exist integers N and k such that for all $n > N$, $n \in S$ if and only if $k|n$. Show that any nonempty set of positive integers that is closed under addition is eventually linear.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Definite Integral Involving a Parameter

E 2910 [1981, 705]. *Proposed by Tian Jing Huang, Szechwan University, People's Republic of China.*

Prove that the function

$$A(h) \equiv \int_0^k\left[1-(h-\cos x)^2\right]^{1/2}\left[4(h-\cos x)^2-1\right]dx,\ k = \arccos(h-1),$$

has at least two zeros in the interval $(0,2)$.

Solution by William A. Newcomb, Lawrence Livermore National Laboratory. We have first

$$A(0) = \int_0^{\pi}\sin x(4\cos^2 x - 1)\,dx = \frac{2}{3} > 0.$$

We have also $A(h) > 0$ for $\frac{3}{2} < h < 2$, since the integrand in this case is positive throughout the range of integration. Now let $w = 1 - 2\cos x$. Then, by straightforward calculation,

$$A(1) = \int_{-1}^{1} F(w)\,dw,$$

where $F(w) = \frac{1}{2}w(2+w)\sqrt{\dfrac{(3+w)(1-w)}{(3-w)(1+w)}}$. Finally, since

$$F(w) + F(-w) = -\frac{w^2(1+w^2)}{\sqrt{(9-w^2)(1-w^2)}},$$

we see that $A(1) < 0$. But $A(h)$ is continuous and must therefore have at least one zero in each of the intervals $(0,1)$ and $(1,3/2)$.

Gaitley found $h = 0.2766, 1.2547$.

Also solved by B. J. Gaitley, D. Hensley, G. A. Heuer, O. P. Lossers (Netherlands), M. Woltermann, and the proposer.

E 3016. *Proposed by Eugene Levine, Adelphi University.*

Let $a_1, a_2, \ldots, a_n$ be positive numbers ($n > 2$), let $s = a_1 + a_2 + \cdots + a_n$, and let $0 < \beta \leqslant 1$. Prove that

$$\sum_{k=1}^{n}\left(\frac{s-a_k}{a_k}\right)^{\beta} \geqslant (n-1)^{2\beta}\sum_{k=1}^{n}\left(\frac{a_k}{s-a_k}\right)^{\beta}.$$

Further, show that equality holds iff $a_1 = a_2 = \cdots = a_n$.

E 3017. *Proposed by Charles W. Schelin, University of Wisconsin, La Crosse.*

Let $f(z)$ be a real polynomial of degree $n \geqslant 1$ such that $f(-1)f(1) \neq 0$. Put $L = -f'(-1)/f(-1)$ and $R = f'(1)/f(1)$. Show that:

(a) If $L \leqslant n/2$ or $R \leqslant n/2$, then f has a zero in $|z| \geqslant 1$.

(b) If $L \geqslant n/2$ or $R \geqslant n/2$, then f has a zero in $|z| \leqslant 1$.

E 3018. *Proposed by A. M. Nadel, Pomona College.*

A nonempty set S of positive integers is said to be eventually linear if there exist integers N and k such that for all $n > N$, $n \in S$ if and only if $k|n$. Show that any nonempty set of positive integers that is closed under addition is eventually linear.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Definite Integral Involving a Parameter

E 2910 [1981, 705]. *Proposed by Tian Jing Huang, Szechwan University, People's Republic of China.*

Prove that the function

$$A(h) \equiv \int_0^k \left[1-(h-\cos x)^2\right]^{1/2}\left[4(h-\cos x)^2-1\right]dx,\ k = \arccos(h-1),$$

has at least two zeros in the interval $(0,2)$.

Solution by William A. Newcomb, Lawrence Livermore National Laboratory. We have first

$$A(0) = \int_0^{\pi}\sin x(4\cos^2 x - 1)\,dx = \frac{2}{3} > 0.$$

We have also $A(h) > 0$ for $\frac{3}{2} < h < 2$, since the integrand in this case is positive throughout the range of integration. Now let $w = 1 - 2\cos x$. Then, by straightforward calculation,

$$A(1) = \int_{-1}^{1} F(w)\,dw,$$

where $F(w) = \frac{1}{2}w(2+w)\sqrt{\dfrac{(3+w)(1-w)}{(3-w)(1+w)}}$. Finally, since

$$F(w) + F(-w) = -\frac{w^2(1+w^2)}{\sqrt{(9-w^2)(1-w^2)}},$$

we see that $A(1) < 0$. But $A(h)$ is continuous and must therefore have at least one zero in each of the intervals $(0,1)$ and $(1,3/2)$.

Gaitley found $h = 0.2766, 1.2547$.

Also solved by B. J. Gaitley, D. Hensley, G. A. Heuer, O. P. Lossers (Netherlands), M. Woltermann, and the proposer.

Numerators and Denominators of Bernoulli Numbers

E 2912 [1981, 705]. *Proposed by Barry Powell, Kirkland, WA.*

Let N_{2m}, D_{2m}, be the numerator and denominator of the $2m$th Bernoulli number B_{2m}, defined by $\tan z = \Sigma(-1)^{n-1}2^{2n}(2^{2n-1}-1)B_{2n}z^{2n-1}/(2n)!$.

Prove: (a) If p is an odd prime, $N_{2p}/p \equiv 1 \bmod p$. (b)* For any even positive integer k, there exist infinitely many even positive integers $r, s, t, \ldots$ such that $N_k|N_r$, $N_k|N_s, \ldots$.

Solution to part (a) by Ira Gessel, Massachusetts Institute of Technology. Note that the assertion is false for $p=3$, since $B_6=\frac{1}{42}$. We prove it for odd primes greater than 3.

By von Staudt's theorem (see, e.g., G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Fourth Edition, p. 91) D_{2m} is the product of all primes q such that $q-1$ divides $2m$. Thus D_{2p} is either 6 or $6(2p+1)$. In either case $D_{2p}\equiv 6(\bmod p)$. By Kummer's congruence (J. Reine Angew. Math., 41 (1851), 368–372), if $p-1$ does not divide $2m$ and if $2m>p$, then

$$B_{2m}/(2m) \equiv B_{2m-p+1}/(2m-p+1)(\bmod p).$$

Thus

$$B_{2p}/(2p) \equiv B_2/2 = 1/12(\bmod p),$$

so

$$N_{2p}/p = \left(B_{2p}/(2p)\right)\cdot 2D_{2p} \equiv (1/12)\cdot 2\cdot 6 \equiv 1(\bmod p).$$

It may be noted that the definition given for the Bernoulli numbers is incorrect: the formula should be

$$\tan z = \sum(-1)^{n-1}2^{2n}(2^{2n}-1)B_{2n}\frac{z^{2n-1}}{(2n)!}.$$

The Bernoulli numbers are usually defined by the much simpler formula

$$z/(e^z-1) = \sum_{n=0}^{\infty} B_n z^n/n!.$$

Solution to part (b) *by L. E. Mattics, University of Southern Alabama.* To answer (b) it suffices to show that for $N_k \geqslant 5$, $N_k|N_r$ where $2|k$ and $r=\phi(N_k^2)+k$. Indeed, with von Staudt's theorem and the formula $B_n(x)=\Sigma_{i=0}^n x^{n-1}\binom{n}{i}B_i$ for Bernoulli polynomials, it is a number-theoretic exercise to deduce that

$$(1)\qquad B_{r+1}(N_k)/(r+1) \equiv B_r N_k\left(\bmod N_k^2\right)$$

$$(2)\qquad B_{k+1}(N_k)/(k+1) \equiv B_k N_k\left(\bmod N_k^2\right) \equiv 0\left(\bmod N_k^2\right)$$

and if $p^s\|N_k$ with p prime and $s>1$,

$$(3)\qquad B_{k+1}(N_k/p) \equiv 0\left(\bmod p^{2s-2}\right).$$

Now if p is prime, $p^s\|N_k$, $s\geqslant 1$, then

$$\begin{aligned}\sum_{i=1}^{N_k-1} i^r &\equiv \sum_{\substack{i=1\\(i,p)=1}}^{N_k-1} i^r \text{ since } r\geqslant 2s\\ &\equiv \sum_{\substack{i=1\\(i,p)=1}}^{N_k-1} i^k \text{ since } r=k+\phi\left(N_k^2\right)\end{aligned}$$

$$\equiv \sum_{\substack{i=1\\(i,p)=1}}^{N_k-1} i^k + \sum_{i=1}^{(N_k/p)-1} p^k i^k \text{ by (3)}$$

$$\equiv \sum_{i=1}^{N_k-1} i^k \pmod{p^{2s}}.$$

Hence by (1) and (2),

$$B_r N_k \equiv B_k N_k \equiv 0 \bmod N_k^2 \text{ and we are done.}$$

Part (a) was also solved by L. E. Mattics, the proposer and by S. Wagstaff, who forwarded important comments.

$3^n = 1 + 2^a + 2^b + \cdots$

E 2915 [1981, 763]. *Proposed by Leo J. Alex, SUNY at Oneonta, NY.*

Let k be fixed. It is conjectured that $3^n = \sum_{i=1}^{k} 2^{r_i}$ has only finitely many integral solutions $(n, r_1, r_2, \ldots, r_k)$. This is clear if $k = 1$; Størmer proved it if $k = 2$, and S. S. Pillai if $k = 3$. Prove the assertion for $k = 4$.

Remark by E. G. Straus, University of California at Los Angeles. The "conjecture" is a corollary of a Theorem of H. G. Senge and mine [Theorem 3, *PV-Numbers and Sets of Multiplicity*, Periodica Math. Hungarica, 3 (1973) 93–100] which shows that, if there are infinitely many integers that can be expressed in two different integral bases b_1, b_2 with sums of digits below a fixed bound k, then b_1 and b_2 are powers of the same integer b.

Also solved by M. Bencze (Romania), L. L. Foster, L. E. Mattics, St. Olaf Problem Group, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by February 29, 1984. The solver's full post-office address should be on each sheet.

6439. *Proposed by Morton Brown, University of Michigan.*

Let $\{a_n\}$ be a sequence of real numbers satisfying the relation $a_{n+1} = |a_n| - a_{n-1}$. Prove that $\{a_n\}$ is periodic with period 9.

6440. *Proposed by M. S. Klamkin, J. McGregor and A. Meir, University of Alberta.*

Let $F(x)$, $G(x)$ be two functions in $L_1(-\infty, \infty)$ which satisfy

$$\int_{-\infty}^{\infty} F(x)\,dx = \int_{-\infty}^{\infty} G(x)\,dx = 1.$$

Show that for any λ in $(0, 1)$ there is a set $E \subseteq (-\infty, \infty)$ such that

$$\int_E F(x)\,dx = \int_E G(x)\,dx = \lambda.$$

6441. *Proposed by F. W. Schmidt and R. Simion, Southern Illinois University.*

Let $a_1 < a_2 < \cdots < a_n$ be n distinct positive integers. Show that the rational function

$$\prod_{1 \leqslant i < j \leqslant n} \frac{x^{a_i} - x^{a_j}}{x^i - x^j}$$

is actually a polynomial.

$$\equiv \sum_{\substack{i=1\\(i,p)=1}}^{N_k-1} i^k + \sum_{i=1}^{(N_k/p)-1} p^k i^k \text{ by (3)}$$

$$\equiv \sum_{i=1}^{N_k-1} i^k \pmod{p^{2s}}.$$

Hence by (1) and (2),

$$B_r N_k \equiv B_k N_k \equiv 0 \bmod N_k^2 \text{ and we are done.}$$

Part (a) was also solved by L. E. Mattics, the proposer and by S. Wagstaff, who forwarded important comments.

$3^n = 1 + 2^a + 2^b + \cdots$

E 2915 [1981, 763]. *Proposed by Leo J. Alex, SUNY at Oneonta, NY.*

Let k be fixed. It is conjectured that $3^n = \sum_{i=1}^{k} 2^{r_i}$ has only finitely many integral solutions $(n, r_1, r_2, \ldots, r_k)$. This is clear if $k = 1$; Størmer proved it if $k = 2$, and S. S. Pillai if $k = 3$. Prove the assertion for $k = 4$.

Remark by E. G. Straus, University of California at Los Angeles. The "conjecture" is a corollary of a Theorem of H. G. Senge and mine [Theorem 3, *PV-Numbers and Sets of Multiplicity*, Periodica Math. Hungarica, 3 (1973) 93–100] which shows that, if there are infinitely many integers that can be expressed in two different integral bases b_1, b_2 with sums of digits below a fixed bound k, then b_1 and b_2 are powers of the same integer b.

Also solved by M. Bencze (Romania), L. L. Foster, L. E. Mattics, St. Olaf Problem Group, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by February 29, 1984. The solver's full post-office address should be on each sheet.

6439. *Proposed by Morton Brown, University of Michigan.*

Let $\{a_n\}$ be a sequence of real numbers satisfying the relation $a_{n+1} = |a_n| - a_{n-1}$. Prove that $\{a_n\}$ is periodic with period 9.

6440. *Proposed by M. S. Klamkin, J. McGregor and A. Meir, University of Alberta.*

Let $F(x)$, $G(x)$ be two functions in $L_1(-\infty, \infty)$ which satisfy

$$\int_{-\infty}^{\infty} F(x)\,dx = \int_{-\infty}^{\infty} G(x)\,dx = 1.$$

Show that for any λ in (0, 1) there is a set $E \subseteq (-\infty, \infty)$ such that

$$\int_E F(x)\,dx = \int_E G(x)\,dx = \lambda.$$

6441. *Proposed by F. W. Schmidt and R. Simion, Southern Illinois University.*

Let $a_1 < a_2 < \cdots < a_n$ be n distinct positive integers. Show that the rational function

$$\prod_{1 \leqslant i < j \leqslant n} \frac{x^{a_i} - x^{a_j}}{x^i - x^j}$$

is actually a polynomial.

$$\equiv \sum_{\substack{i=1\\(i,p)=1}}^{N_k-1} i^k + \sum_{i=1}^{(N_k/p)-1} p^k i^k \text{ by (3)}$$

$$\equiv \sum_{i=1}^{N_k-1} i^k \pmod{p^{2s}}.$$

Hence by (1) and (2),

$$B_r N_k \equiv B_k N_k \equiv 0 \bmod N_k^2 \text{ and we are done.}$$

Part (a) was also solved by L. E. Mattics, the proposer and by S. Wagstaff, who forwarded important comments.

$3^n = 1 + 2^a + 2^b + \cdots$

E 2915 [1981, 763]. *Proposed by Leo J. Alex, SUNY at Oneonta, NY.*

Let k be fixed. It is conjectured that $3^n = \sum_{i=1}^{k} 2^{r_i}$ has only finitely many integral solutions $(n, r_1, r_2, \dots, r_k)$. This is clear if $k = 1$; Størmer proved it if $k = 2$, and S. S. Pillai if $k = 3$. Prove the assertion for $k = 4$.

Remark by E. G. Straus, University of California at Los Angeles. The "conjecture" is a corollary of a Theorem of H. G. Senge and mine [Theorem 3, *PV-Numbers and Sets of Multiplicity*, Periodica Math. Hungarica, 3 (1973) 93–100] which shows that, if there are infinitely many integers that can be expressed in two different integral bases b_1, b_2 with sums of digits below a fixed bound k, then b_1 and b_2 are powers of the same integer b.

Also solved by M. Bencze (Romania), L. L. Foster, L. E. Mattics, St. Olaf Problem Group, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by February 29, 1984. The solver's full post-office address should be on each sheet.

6439. *Proposed by Morton Brown, University of Michigan.*

Let $\{a_n\}$ be a sequence of real numbers satisfying the relation $a_{n+1} = |a_n| - a_{n-1}$. Prove that $\{a_n\}$ is periodic with period 9.

6440. *Proposed by M. S. Klamkin, J. McGregor and A. Meir, University of Alberta.*

Let $F(x)$, $G(x)$ be two functions in $L_1(-\infty, \infty)$ which satisfy

$$\int_{-\infty}^{\infty} F(x)\,dx = \int_{-\infty}^{\infty} G(x)\,dx = 1.$$

Show that for any λ in $(0, 1)$ there is a set $E \subseteq (-\infty, \infty)$ such that

$$\int_E F(x)\,dx = \int_E G(x)\,dx = \lambda.$$

6441. *Proposed by F. W. Schmidt and R. Simion, Southern Illinois University.*

Let $a_1 < a_2 < \cdots < a_n$ be n distinct positive integers. Show that the rational function

$$\prod_{1 \leqslant i < j \leqslant n} \frac{x^{a_i} - x^{a_j}}{x^i - x^j}$$

is actually a polynomial.

SOLUTIONS OF ADVANCED PROBLEMS

Flexible Linkages of Polygons

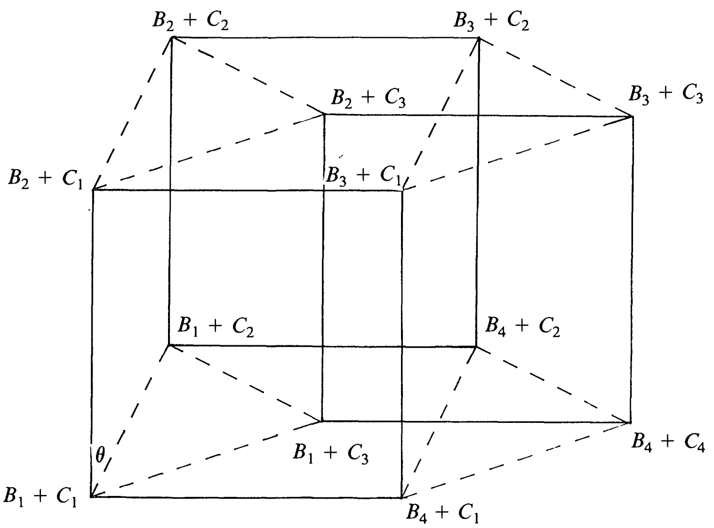
6367 [1981, 711]. *Proposed by A. Ehrenfeucht and J. Mycielski, University of Colorado.*

Let A be a finite collection of distinct but possibly overlapping regular n-gons of the same size on the plane such that every vertex of every n-gon of A is a vertex of exactly two n-gons of A.

(a) Construct a collection A of $2n$ n-gons such that, even if the n-gons are rigid, A is flexible.

(b)* For which n is a rigid A possible?

Solution to part (a) *by Kit Hanes, Eastern Washington University.* We establish a more general result. Let B be any n-gon with vertices $B_1,\ldots,B_n$ and let C be any m-gon with vertices $C_1,\ldots,C_m$. Let θ be the angle between $\overline{B_1B_2}$ and $\overline{C_1C_2}$. Then, for $i \in \{1,\ldots,n\}$ let $\hat{C}_i = B_i + C$, and for $j \in \{1,\ldots,m\}$ let $\hat{B}_j = B + C_j$. Let $A = \{\hat{C}_1,\ldots,\hat{C}_n, \hat{B}_1,\ldots,\hat{B}_m\}$. A is a collection of $n+m$ polygons each of which is a translate of either B or C. Any vertex of any polygon in A is a point $B_i + C_j$, which is a vertex just of $\hat{B}_j$ and of $\hat{C}_i$. This construction is valid independently of θ; as θ varies A flexes.



Also solved by Michael Goldberg and the proposers.

Part (b). The proposers state that they know of only two cases when A as defined in the problem is rigid, namely $n = 2$ (a triangle) and $n = 4$ (a configuration of 12 squares inside a regular 12-gon whose outer edges form the 12-gon).

ANSWER TO PHOTOS ON PAGE 517

They are all outstanding mathematicians, and they were all presidents of the A.M.S. Here are their names, followed by the years of their presidency. Top row: J. von Neumann (1951–1952), R. L. Wilder (1955–1956); bottom row: J. L. Doob (1963–1964), Oscar Zariski (1969–1970).

A Sum of Reciprocals of Euler Totients

6370* [1981, 769]. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

For each set of positive integers $\{a_i\}$ such that $\Sigma a_i^{-1} \leqslant 1$, define $f_1(n) := \Sigma 1/\phi(a_i)$, the sum being extended over those $a_i \leqslant n$. Set $f(n) = \max f_1(n)$, the maximum being taken over the admissible sets $\{a_i\}$; ϕ is Euler's totient. Estimate $f(n)$ as well as you can.

Solution by M. G. Hickman, New Mexico State University. Let

$$a = p_1 p_2 \cdots p_k \leqslant n < p_1 p_2 \cdots p_{k+1}$$

where $p_1, p_2, \ldots$ are the primes in increasing order. Then

$$f_1(n) \leqslant \max_{m \leqslant n} \frac{m}{\phi(m)}.$$

Further, if $m \leqslant n$ and the distinct prime factors of m are $q_1, q_2, \ldots, q_j$, then $j \leqslant k$ and hence

$$\frac{\phi(m)}{m} = \prod_{i=1}^{j}\left(1 - \frac{1}{q_i}\right) \geqslant \prod_{i=1}^{k}\left(1 - \frac{1}{p_i}\right) = \frac{\phi(a)}{a}.$$

It follows that

$$f(n) \leqslant \frac{a}{\phi(a)}.$$

If repetitions of the integers a_i are allowed in the definition of $f_1(n)$, we obtain

$$f(n) = \frac{a}{\phi(a)}$$

by taking $a_i = a$ for $i = 1, 2, \ldots, a$ and observing that, in this case, $f_1(n) = a/\phi(a)$. If repetitions are not permitted, the above yields only an upper bound on $f(n)$.

Closed and Connected Graphs

6373 [1981, 769]. *Proposed by F. S. Cater, Portland State University.*

By an "order space" we mean a totally ordered nonvoid set endowed with the open interval topology. Let X be an order space, and let Y be either an order space or a locally compact Hausdorff space. Let f be a mapping of X into Y that is not everywhere continuous, and let G denote the graph of f, $G = \{(x, f(x)) : x \in X\}$.

(i) Prove that G is not both a closed and connected subset of the product space $X \times Y$. (Compare problem 6255 [1980, 679] in this MONTHLY.)

(ii) Show, by example, that (i) need not hold for complete connected metric spaces X and $Y = f(X)$, if X is not an order space.

Solution to part (i) by the proposer. Suppose that G is closed and connected and f is discontinuous at x. There is a neighbourhood V of $f(x)$ such that for any neighbourhood U of x, there is an $x' \in U$ with $f(x') \notin V$. By reducing the neighbourhood V of $f(x)$, if necessary, we can (and do) assume that bdry V is compact. (Note here that, if Y is an order space, intervals have boundaries consisting of at most two points.)

We claim that there is a neighbourhood U_0 of x such that $f(U_0) \subset V \cup \operatorname{ext} V$. Assume not. For each neighbourhood U of x, pick $x_U \in U$ with $f(x_U) \in$ bdry V. Let $\mathcal{U}$ denote the set of neighbourhoods of x directed by inclusion. Then $\{f(y_U)\}$ is a net on this directed set, and a subnet of $\{f(x_U)\}$ converges to y, say, in the compact set, bdry V. It follows that the point (x, y) is in the closure of G. But $(x, f(x)) \in G$ and $y \neq f(x)$. This contradicts the hypothesis that f is a function and $G = \overline{G}$.

A Sum of Reciprocals of Euler Totients

6370* [1981, 769]. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

For each set of positive integers $\{a_i\}$ such that $\Sigma a_i^{-1} \leqslant 1$, define $f_1(n){:} = \Sigma 1/\phi(a_i)$, the sum being extended over those $a_i \leqslant n$. Set $f(n) = \max f_1(n)$, the maximum being taken over the admissible sets $\{a_i\}$; ϕ is Euler's totient. Estimate $f(n)$ as well as you can.

Solution by M. G. Hickman, New Mexico State University. Let

$$a = p_1 p_2 \cdots p_k \leqslant n < p_1 p_2 \cdots p_{k+1}$$

where $p_1, p_2, \ldots$ are the primes in increasing order. Then

$$f_1(n) \leqslant \max_{m \leqslant n} \frac{m}{\phi(m)}.$$

Further, if $m \leqslant n$ and the distinct prime factors of m are $q_1, q_2, \ldots, q_j$, then $j \leqslant k$ and hence

$$\frac{\phi(m)}{m} = \prod_{i=1}^{j}\left(1 - \frac{1}{q_i}\right) \geqslant \prod_{i=1}^{k}\left(1 - \frac{1}{p_i}\right) = \frac{\phi(a)}{a}.$$

It follows that

$$f(n) \leqslant \frac{a}{\phi(a)}.$$

If repetitions of the integers a_i are allowed in the definition of $f_1(n)$, we obtain

$$f(n) = \frac{a}{\phi(a)}$$

by taking $a_i = a$ for $i = 1, 2, \ldots, a$ and observing that, in this case, $f_1(n) = a/\phi(a)$. If repetitions are not permitted, the above yields only an upper bound on $f(n)$.

Closed and Connected Graphs

6373 [1981, 769]. *Proposed by F. S. Cater, Portland State University.*

By an "order space" we mean a totally ordered nonvoid set endowed with the open interval topology. Let X be an order space, and let Y be either an order space or a locally compact Hausdorff space. Let f be a mapping of X into Y that is not everywhere continuous, and let G denote the graph of f, $G = \{(x, f(x)) : x \in X\}$.

(i) Prove that G is not both a closed and connected subset of the product space $X \times Y$. (Compare problem 6255 [1980, 679] in this MONTHLY.)

(ii) Show, by example, that (i) need not hold for complete connected metric spaces X and $Y = f(X)$, if X is not an order space.

Solution to part (*i*) *by the proposer.* Suppose that G is closed and connected and f is discontinuous at x. There is a neighbourhood V of $f(x)$ such that for any neighbourhood U of x, there is an $x' \in U$ with $f(x') \notin V$. By reducing the neighbourhood V of $f(x)$, if necessary, we can (and do) assume that bdry V is compact. (Note here that, if Y is an order space, intervals have boundaries consisting of at most two points.)

We claim that there is a neighbourhood U_0 of x such that $f(U_0) \subset V \cup \operatorname{ext} V$. Assume not. For each neighbourhood U of x, pick $x_U \in U$ with $f(x_U) \in$ bdry V. Let $\mathcal{U}$ denote the set of neighbourhoods of x directed by inclusion. Then $\{f(y_U)\}$ is a net on this directed set, and a subnet of $\{f(x_U)\}$ converges to y, say, in the compact set, bdry V. It follows that the point (x, y) is in the closure of G. But $(x, f(x)) \in G$ and $y \neq f(x)$. This contradicts the hypothesis that f is a function and $G = \overline{G}$.

We reduce U_0 to an interval $[a,b]$, where $x \in \operatorname{int}[a,b]$, $f[a,b] \subset V \cup \operatorname{ext} V$. But some $c \in \operatorname{int}[a,b]$ satisfies $f(c) \notin V$. Say $c < x$. (The case $c > x$ is analogous.) Then $f(x) \in V$, $f(c) \in \operatorname{ext} V$, $f[c,x] \subset V \cup \operatorname{ext} V$. The following two sets are disjoint and open in $X \times Y$:

$$((c,x] \times V) \cup ((x,\infty) \times Y), \quad ([c,x) \times \operatorname{ext} V) \cup ((-\infty,c) \times Y).$$

The union of these sets contains the graph G. The first set contains the point $(x, f(x))$ and the second contains $(c, f(c))$. Thus G is not connected, contrary to hypothesis.

Solution to part (ii) by D. E. Sanderson, Iowa State University. Let X be the unit circle and in polar coordinates set $f(1,\theta) = 2\pi/(2\pi - \theta)$. Then $Y = f(X) = [1,\infty)$ and f is discontinuous at $(1,0)$, yet its graph is closed and connected.

A Sequence Converging to π

6375 [1982, 65]. *Proposed by Chang Gengzhe, University of Utah.*

Set

$$1 + \frac{n-1}{n+2} + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right) + \cdots + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right) \cdots \left(\frac{1}{2n}\right).$$

Show that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} S_n = \frac{\sqrt{\pi}}{2}.$$

Composite solution. We have

$$S_n = \binom{2n}{n-1}^{-1} \sum_{k=0}^{n-1} \binom{2n}{k}$$

$$= \frac{1}{2}\binom{2n}{n-1}^{-1}\left(2^{2n} - \binom{2n}{n}\right) = 2^{2n-1}\binom{2n}{n-1}^{-1} - \frac{n+1}{2n}.$$

Application of Stirling's approximation formula to the binomial coefficient yields the desired result.

Solved by 52 solvers including the proposer. Some solvers used the Wallis formula

$$\pi = \lim_{n\to\infty} \frac{1}{n} \prod_{i=1}^{n} \left(1 + \frac{1}{2i-1}\right)^2$$

instead of Stirling's approximation formula.

L. Van Hamme (Belgium) obtained the more general result

$$\frac{S_n}{\sqrt{n}} = \frac{\sqrt{\pi}}{2} - \frac{1}{2\sqrt{n}} + \frac{9\sqrt{\pi}}{16n} - \frac{1}{2n^{3/2}} + \theta\left(\frac{1}{n^2}\right).$$

G. H. Gonnet showed that if r and s are arbitrary constants and

$$S_n(r,s) = 1 + \frac{n+r}{n+s} + \frac{(n+r)(n+r-1)}{(n+s)(n+s+1)} + \cdots + \frac{\Gamma(n+s)\Gamma(n+r+1)}{\Gamma(n+s+k)\Gamma(n+r+1-k)} + \cdots,$$

then

$$\frac{S_n(r,s)}{\sqrt{n}} = \frac{\sqrt{\pi}}{2} + \frac{2+r-s}{2\sqrt{n}} + \theta\left(\frac{1}{n}\right).$$

We reduce U_0 to an interval $[a,b]$, where $x \in \operatorname{int}[a,b]$, $f[a,b] \subset V \cup \operatorname{ext} V$. But some $c \in \operatorname{int}[a,b]$ satisfies $f(c) \notin V$. Say $c < x$. (The case $c > x$ is analogous.) Then $f(x) \in V$, $f(c) \in \operatorname{ext} V$, $f[c,x] \subset V \cup \operatorname{ext} V$. The following two sets are disjoint and open in $X \times Y$:

$$((c,x] \times V) \cup ((x,\infty) \times Y), \quad ([c,x) \times \operatorname{ext} V) \cup ((-\infty,c) \times Y).$$

The union of these sets contains the graph G. The first set contains the point $(x, f(x))$ and the second contains $(c, f(c))$. Thus G is not connected, contrary to hypothesis.

Solution to part (ii) by D. E. Sanderson, Iowa State University. Let X be the unit circle and in polar coordinates set $f(1,\theta) = 2\pi/(2\pi - \theta)$. Then $Y = f(X) = [1,\infty)$ and f is discontinuous at $(1,0)$, yet its graph is closed and connected.

A Sequence Converging to π

6375 [1982, 65]. *Proposed by Chang Gengzhe, University of Utah.*

Set

$$1 + \frac{n-1}{n+2} + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right) + \cdots + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right)\cdots\left(\frac{1}{2n}\right).$$

Show that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} S_n = \frac{\sqrt{\pi}}{2}.$$

Composite solution. We have

$$S_n = \binom{2n}{n-1}^{-1} \sum_{k=0}^{n-1} \binom{2n}{k}$$

$$= \frac{1}{2}\binom{2n}{n-1}^{-1}\left(2^{2n} - \binom{2n}{n}\right) = 2^{2n-1}\binom{2n}{n-1}^{-1} - \frac{n+1}{2n}.$$

Application of Stirling's approximation formula to the binomial coefficient yields the desired result.

Solved by 52 solvers including the proposer. Some solvers used the Wallis formula

$$\pi = \lim_{n\to\infty} \frac{1}{n} \prod_{i=1}^{n} \left(1 + \frac{1}{2i-1}\right)^2$$

instead of Stirling's approximation formula.

L. Van Hamme (Belgium) obtained the more general result

$$\frac{S_n}{\sqrt{n}} = \frac{\sqrt{\pi}}{2} - \frac{1}{2\sqrt{n}} + \frac{9\sqrt{\pi}}{16n} - \frac{1}{2n^{3/2}} + \theta\left(\frac{1}{n^2}\right).$$

G. H. Gonnet showed that if r and s are arbitrary constants and

$$S_n(r,s) = 1 + \frac{n+r}{n+s} + \frac{(n+r)(n+r-1)}{(n+s)(n+s+1)} + \cdots + \frac{\Gamma(n+s)\Gamma(n+r+1)}{\Gamma(n+s+k)\Gamma(n+r+1-k)} + \cdots,$$

then

$$\frac{S_n(r,s)}{\sqrt{n}} = \frac{\sqrt{\pi}}{2} + \frac{2+r-s}{2\sqrt{n}} + \theta\left(\frac{1}{n}\right).$$

Limit of a Sequence

6376 [1982, 65]. *Proposed by I. J. Schoenberg, University of Wisconsin.*

Let a_n ($n = 1,2,\dots$) be reals. Show that

$$\lim_{n\to\infty} a_n \sum_1^n a_i^2 = 1 \text{ implies that } \lim_{n\to\infty} (3n)^{1/3} a_n = 1.$$

Solution by A. Meir, University of Alberta. Let $s_n = \sum_{i=1}^n a_i^2$. Then the condition $a_n s_n \to 1$ implies that $s_n \to \infty$ and $a_n \to 0$. Hence we also have that $a_n s_{n-1} \to 1$. Therefore

$$s_n^3 - s_{n-1}^3 = a_n^2\left(s_n^2 + s_n s_{n-1} + s_{n-1}^2\right) \to 3$$

and thus $s_n^3 \sim 3n$. It follows that $a_n \sim (3n)^{-1/3}$.

Similarly, it can be proved that if for some $p > 0$,

$$\lim_{n\to\infty} a_n \sum_{i=1}^n |a_i|^p = 1,$$

then $a_n \sim ((p+1)n)^{-1/(p+1)}$.

Also solved by E. J. Barbeau (Canada), Artin Boghossian (Saudi Arabia), Robert Breusch, Paul S. Bruckman, Edmond Butler, Jeesen Chen, David C. Cox, Nathaniel Grossman, Ellen Hertz, Robert B. Israel (Canada), Mauri Koskela (Finland), T. K. Louton & C. C. Rousseau, L. E. Mattics, Robert K. Meany, William A. Newcomb, Johannes C. C. Nitsche, Otto G. Ruehr, Chow Seong Seow (Canada), W. T. Sledd, and the proposer.

The proposer wishes to add that he received the problem without solution from Franklin Richards about ten years ago.

Decomposition of a Function into Measurable Functions

6378 [1982, 134]. *Proposed by Yi Hong and Jingcheng Tong, Wayne State University.*

Let $f(x)$ be a real function defined on $I = [0,1]$. Prove there are two Lebesgue measurable functions $g(x)$ and $h(x)$ such that $f(x) = g(h(x))$.

Solution. Berthold Schweizer and John C. Morgan II pointed out that the result is known, apparently having been obtained first by S. Ruziewicz in 1932 and subsequently published in his paper, "Sur une propriété des fonctions arbitraires d'une variable réelle," Mathematica, 9 (1935) 83–85.

Michael Renardy and the proposers submitted the following decomposition of f. Define h: $I \to I$ by setting

$$h(x) = \sum_{n=1}^\infty \frac{2a_n}{3^n}$$

when x has the nonterminating binary representation

$$x = \sum_{n=1}^\infty \frac{a_n}{2^n}, \quad a_n \in \{0,1\}.$$

Then h is strictly increasing on I, and $h(I)$, being a subset of the Cantor set, has Lebesgue measure zero. Define g: $I \to I$ by setting

$$g(x) = \begin{cases} f\left(h^{-1}(x)\right) & \text{for } x \in h(I), \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x) = g(h(x))$ for $x \in I$, and both g and h are Lebesgue measurable.

Limit of a Sequence

6376 [1982, 65]. *Proposed by I. J. Schoenberg, University of Wisconsin.*

Let a_n $(n = 1, 2, \ldots)$ be reals. Show that

$$\lim_{n\to\infty} a_n \sum_1^n a_i^2 = 1 \text{ implies that } \lim_{n\to\infty} (3n)^{1/3} a_n = 1.$$

Solution by A. Meir, University of Alberta. Let $s_n = \sum_{i=1}^n a_i^2$. Then the condition $a_n s_n \to 1$ implies that $s_n \to \infty$ and $a_n \to 0$. Hence we also have that $a_n s_{n-1} \to 1$. Therefore

$$s_n^3 - s_{n-1}^3 = a_n^2\left(s_n^2 + s_n s_{n-1} + s_{n-1}^2\right) \to 3$$

and thus $s_n^3 \sim 3n$. It follows that $a_n \sim (3n)^{-1/3}$.

Similarly, it can be proved that if for some $p > 0$,

$$\lim_{n\to\infty} a_n \sum_{i=1}^n |a_i|^p = 1,$$

then $a_n \sim ((p+1)n)^{-1/(p+1)}$.

Also solved by E. J. Barbeau (Canada), Artin Boghossian (Saudi Arabia), Robert Breusch, Paul S. Bruckman, Edmond Butler, Jeesen Chen, David C. Cox, Nathaniel Grossman, Ellen Hertz, Robert B. Israel (Canada), Mauri Koskela (Finland), T. K. Louton & C. C. Rousseau, L. E. Mattics, Robert K. Meany, William A. Newcomb, Johannes C. C. Nitsche, Otto G. Ruehr, Chow Seong Seow (Canada), W. T. Sledd, and the proposer.

The proposer wishes to add that he received the problem without solution from Franklin Richards about ten years ago.

Decomposition of a Function into Measurable Functions

6378 [1982, 134]. *Proposed by Yi Hong and Jingcheng Tong, Wayne State University.*

Let $f(x)$ be a real function defined on $I = [0, 1]$. Prove there are two Lebesgue measurable functions $g(x)$ and $h(x)$ such that $f(x) = g(h(x))$.

Solution. Berthold Schweizer and John C. Morgan II pointed out that the result is known, apparently having been obtained first by S. Ruziewicz in 1932 and subsequently published in his paper, "Sur une propriété des fonctions arbitraires d'une variable réelle," Mathematica, 9 (1935) 83–85.

Michael Renardy and the proposers submitted the following decomposition of f. Define h: $I \to I$ by setting

$$h(x) = \sum_{n=1}^{\infty} \frac{2a_n}{3^n}$$

when x has the nonterminating binary representation

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}, \quad a_n \in \{0, 1\}.$$

Then h is strictly increasing on I, and $h(I)$, being a subset of the Cantor set, has Lebesgue measure zero. Define g: $I \to I$ by setting

$$g(x) = \begin{cases} f\left(h^{-1}(x)\right) & \text{for } x \in h(I), \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x) = g(h(x))$ for $x \in I$, and both g and h are Lebesgue measurable.

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

The Mathematical Theory of Chromatic Plane Ornaments. By Thomas W. Wieting. Marcel Dekker, New York, 1982.

MARJORIE SENECHAL
Department of Mathematics, Smith College, Northampton, MA 01063

Browsing through collections of patterns or ornaments, almost everyone becomes fascinated by their intricate repetitions: we correctly sense that, despite their diversity, all the patterns are variations on a few basic themes. Symmetrically colored patterns are especially attractive and intriguing because of the lively but orderly interaction of permutations and symmetries that they display. A careful scrutiny of colored patterns leads us naturally to consider color symmetries, and to the problem of classifying them.

What is a color symmetry? Suppose that we are given a pattern with symmetry group G. If the pattern is symmetrically colored, we observe that through its action every g in G effects a permutation π_g of the k colors. For example a 90° rotation about the center of any square of an infinite checkerboard effects the identity permutation while a 90° rotation about any corner carries with it an interchange of colors. The pair (g, π_g) is then a color symmetry, and the set $\{(g, \pi_g), g \in G\}$ a color group. It is easy to verify that π_g is the image of g in S_k under a permutation representation of G. The color groups, under a suitable definition of equivalence, classify the subthemes on which all colored patterns are based.

Historically, crystallography was the primary motivation for color groups, just as it was the original motivation for the mathematical analysis of repeating patterns and the enumeration of the crystallographic groups. The theory was developed in response to the need of crystal chemistry and crystal physics for a classification scheme for patterns within patterns. This need arises in the study of the geometrical relations among crystal structures, and in the classification of the ways in which magnetic moments can be symmetrically distributed among the atoms in a single structure. To handle these and related applications, we need a complete enumeration of the k-color groups for each k, up to some reasonably large number, in 2- and 3-dimensional space. This is a subtle mathematical problem involving geometry, group theory, linear algebra, and a good deal of computation. There is a richness of material here that can form the basis for a valuable interdisciplinary course for advanced undergraduates.

The cornerstone of such a course must be the theory of crystallographic groups. Bieberbach's 1910 solution of the first part of Hilbert's 18th problem, together with the 1912 discovery of the diffraction of x-rays by crystals, brought these groups into the mainstream of modern mathematics and physics. Although they have rarely found their way into the undergraduate curriculum, they deserve a place there. Not only are they intrinsically interesting, attractive and important; they can also serve as valuable examples in many ways. (In algebra they give us easily visualized examples of group extensions with nontrivial factor sets.) The color groups are a particularly interesting generalization of the crystallographic groups whose heuristic value should also be exploited. For example, by inspecting a colored plane ornament one can quickly write down the permutation representation, find a subgroup which defines it, visualize the action of its cosets, determine whether the subgroup is normal and if it is not, compute the number of its conjugates. The insights obtained in this way can be very valuable, and not only to a beginner.

The book under review thus fills a significant gap. Designed as a text, it is also a definitive monograph, whose goal is to classify and enumerate (for $k \leqslant 60$) the k-color groups of the plane. The material is divided into three parts: (1) the study of the Euclidean plane and its transforma-

tions, (2) the properties and enumeration of the seventeen isomorphism classes of plane crystallographic groups, and (3) the classification and enumeration of their transitive permutation representations. The exposition is careful and clear, with a blend of rigor and informality which is appropriate for the undergraduate level. Several features are especially noteworthy. The relation between the geometric and algebraic properties of the Euclidean plane is studied with unusual care, making the first part an excellent introduction to transformation geometry. In the second part, the two-dimensional crystallographic groups are given the prominent position that the subject requires and which they deserve. A unique and especially valuable feature of the third section is the adaptation of the algebraic conditions for the classification of permutation representations to a set of algorithms suitable for machine computation, and the presentation of extensive, definitive tables of data which correct errors in the recent literature. References to the historical and current mathematical literature can be found in the extensive bibliography; one could wish, however, that it was annotated and that references to applications had been included. Challenging problems at the end of each chapter extend the theory and introduce related topics. The material is a natural introduction to many problems of current interest, including n-dimensional crystallography, tilings and patterns, a deeper study of the geometry of colored patterns, and the algorithmic analysis of classification procedures.

It was disappointing to find that all the illustrations, including the folio of the 96 4-color ornaments, are line drawings: representing colors by circles, dots and dashes does not provide a comparable intuitive grasp of structure. (Imagine, for example, what the impact of Escher's multicolored tessellations on the scientific imagination would have been if he had shaded rather than colored them!) This defect, however, can easily be remedied with a good set of colored pencils. The book is strongly recommended as a thorough introduction to the theory of color groups, as a supplement to various undergraduate courses and above all as the text for a stimulating and significant course.

A Concrete Introduction to Higher Algebra. By Lindsay Childs. Springer-Verlag, New York, 1979. xiv + 388 pp.

MICHAEL ROSEN
Department of Mathematics, Brown University, Providence, RI 02912

Abstract algebra has proved to be so powerful a tool that for a long time even beginning textbooks would emphasize axiomatics and abstraction at the expense of calculation. For example, students would learn the fundamental theorem of Galois theory without learning a thing about computing the Galois group of an explicitly given polynomial. On a more elementary level they would learn that $Q[X]$ is a unique factorization domain but have no idea how to factor a given polynomial, $x^6 + x^5 - 4x^4 - 4x^3 + 2x^2 + 4x + 1$ say, into a product of irreducibles. One of the great virtues of the book under review is that after reading it, one will no longer feel so helpless when confronted with so elementary a task.

This book is very aptly described by its title. By developing the basic ideas of number theory and the theory of polynomials (one variable) the author instructs by means of examples, computations, and fascinating applications. The style is leisurely and informal, a guided tour through the foothills, the guide unable to resist numerous side-paths and return visits to favorite spots. The abstract theories of the higher elevations are only hinted at and occasionally glimpsed.

There is little overlap between the content of this book and that of the more usual text in algebra intended for use at the sophmore-junior level. Groups are mentioned only briefly, but subgroups are omitted, rings are mentioned, but ideals only appear on the last page. Fields receive more attention, but Galois theory not at all. Instead the emphasis is on examples; the integers, the

tions, (2) the properties and enumeration of the seventeen isomorphism classes of plane crystallographic groups, and (3) the classification and enumeration of their transitive permutation representations. The exposition is careful and clear, with a blend of rigor and informality which is appropriate for the undergraduate level. Several features are especially noteworthy. The relation between the geometric and algebraic properties of the Euclidean plane is studied with unusual care, making the first part an excellent introduction to transformation geometry. In the second part, the two-dimensional crystallographic groups are given the prominent position that the subject requires and which they deserve. A unique and especially valuable feature of the third section is the adaptation of the algebraic conditions for the classification of permutation representations to a set of algorithms suitable for machine computation, and the presentation of extensive, definitive tables of data which correct errors in the recent literature. References to the historical and current mathematical literature can be found in the extensive bibliography; one could wish, however, that it was annotated and that references to applications had been included. Challenging problems at the end of each chapter extend the theory and introduce related topics. The material is a natural introduction to many problems of current interest, including n-dimensional crystallography, tilings and patterns, a deeper study of the geometry of colored patterns, and the algorithmic analysis of classification procedures.

It was disappointing to find that all the illustrations, including the folio of the 96 4-color ornaments, are line drawings: representing colors by circles, dots and dashes does not provide a comparable intuitive grasp of structure. (Imagine, for example, what the impact of Escher's multicolored tessellations on the scientific imagination would have been if he had shaded rather than colored them!) This defect, however, can easily be remedied with a good set of colored pencils. The book is strongly recommended as a thorough introduction to the theory of color groups, as a supplement to various undergraduate courses and above all as the text for a stimulating and significant course.

A Concrete Introduction to Higher Algebra. By Lindsay Childs. Springer-Verlag, New York, 1979. xiv + 388 pp.

MICHAEL ROSEN
Department of Mathematics, Brown University, Providence, RI 02912

Abstract algebra has proved to be so powerful a tool that for a long time even beginning textbooks would emphasize axiomatics and abstraction at the expense of calculation. For example, students would learn the fundamental theorem of Galois theory without learning a thing about computing the Galois group of an explicitly given polynomial. On a more elementary level they would learn that $Q[X]$ is a unique factorization domain but have no idea how to factor a given polynomial, $x^6 + x^5 - 4x^4 - 4x^3 + 2x^2 + 4x + 1$ say, into a product of irreducibles. One of the great virtues of the book under review is that after reading it, one will no longer feel so helpless when confronted with so elementary a task.

This book is very aptly described by its title. By developing the basic ideas of number theory and the theory of polynomials (one variable) the author instructs by means of examples, computations, and fascinating applications. The style is leisurely and informal, a guided tour through the foothills, the guide unable to resist numerous side-paths and return visits to favorite spots. The abstract theories of the higher elevations are only hinted at and occasionally glimpsed.

There is little overlap between the content of this book and that of the more usual text in algebra intended for use at the sophmore-junior level. Groups are mentioned only briefly, but subgroups are omitted, rings are mentioned, but ideals only appear on the last page. Fields receive more attention, but Galois theory not at all. Instead the emphasis is on examples; the integers, the

ring of polynomials, the integers modulo m, finite fields, the rational, real, and complex numbers, etc. This approach has the advantage that the material is interesting throughout. There are few dry spots. Gratification, if not immediate, is not long postponed.

To communicate something of the flavor of this text, it is worthwhile to give a sample of some of the many applications which are treated. There are two chapters on secret codes, two chapters on error-correcting codes, a chapter on duplicate bridge tournaments, one on telephone cable splicing, and one on latin squares. It should be mentioned that the chapters, thirty-three in number, are generally quite short. One, which I especially liked, discusses the partial fraction decomposition of rational functions and applies the results to the problem of finding indefinite integrals of rational functions, and also to a problem in the theory of partitions.

As has already been indicated, a lot of space is devoted to factoring polynomials, especially over Q and Z/pZ. Some of the explicit calculations are quite delightful. A pretty application is a proof of the law of quadratic reciprocity due to R. Swan based on a theorem of L. Stickelberger about factoring polynomials over finite fields.

What we have here is a very readable introduction to higher algebra. There is lacking, however, any substantial amount of higher algebra itself. The question arises of how this book can be utilized in the standard undergraduate mathematics curriculum. It is hard to see how it could be used in place of more traditional texts as, for example, those of Birkhoff-Mac Lane, Fraleigh, or Herstein. The author simply does not present enough group, ring, and field theory. There is another problem. The course in abstract algebra is often used to introduce the student to standards of mathematical rigor and proof. The style of Childs' book is too informal to serve this purpose well. An ideal set-up would be a two-semester course, the first out of Child's book, and the second out of a more conventional text. Alternatively, it could be used as the basis of a course for nonmath majors, students who would find the material interesting in itself, and who are not in need of the foundations of abstract algebra as the basis for the study of more advanced mathematics.

Finally, the book is ideal for independent reading. It might be especially suitable for someone who is acquainted with theory but has neglected computations. Calculations and examples not only illuminate theory, they suggest theory. Moreover, mastering a new algorithm is often just plain fun. Perhaps that is why Child's book is not only instructive, it is very entertaining.

Calculus and Analytic Geometry. By C. H. Edwards, Jr., and David E. Penney. Prentice-Hall, Englewood Cliffs, New Jersey, 1982. xvi + 895 pp. $29.95.

PETER ROSENTHAL
Department of Mathematics, University of Toronto, Toronto, Ontario M5S 1A1, Canada

It's spring and our thoughts turn to... choosing a calculus text for next year. We sit quietly, staring at the 4000 pounds of textbooks that aggressive book publishers have sent us. What was wrong with the old Courant? Even the old (pre-fourth) edition of Thomas wasn't so bad.

The sales representative of Prindle, Weber & Schmidt walked into my office and said, "I'm glad to see you're using Swokowski's *Calculus* this year; how do you like it?"

"Nothing personal," I answered, "but it stinks."

"Well," the salesman replied, "It's the best-selling calculus book in North America."

Apparently the salesman was correct; *Calculus with Analytic Geometry* by E. Swokowski is among the best-selling calculus books. I think I was correct too: it stinks. Many of us have apparently been allowing our choice of calculus text to be determined more by the hard-sell techniques of the publishers than by pedagogical considerations.

In selecting calculus texts we get involved in discussions of various points: the level of the text,

ring of polynomials, the integers modulo m, finite fields, the rational, real, and complex numbers, etc. This approach has the advantage that the material is interesting throughout. There are few dry spots. Gratification, if not immediate, is not long postponed.

To communicate something of the flavor of this text, it is worthwhile to give a sample of some of the many applications which are treated. There are two chapters on secret codes, two chapters on error-correcting codes, a chapter on duplicate bridge tournaments, one on telephone cable splicing, and one on latin squares. It should be mentioned that the chapters, thirty-three in number, are generally quite short. One, which I especially liked, discusses the partial fraction decomposition of rational functions and applies the results to the problem of finding indefinite integrals of rational functions, and also to a problem in the theory of partitions.

As has already been indicated, a lot of space is devoted to factoring polynomials, especially over Q and Z/pZ. Some of the explicit calculations are quite delightful. A pretty application is a proof of the law of quadratic reciprocity due to R. Swan based on a theorem of L. Stickelberger about factoring polynomials over finite fields.

What we have here is a very readable introduction to higher algebra. There is lacking, however, any substantial amount of higher algebra itself. The question arises of how this book can be utilized in the standard undergraduate mathematics curriculum. It is hard to see how it could be used in place of more traditional texts as, for example, those of Birkhoff-Mac Lane, Fraleigh, or Herstein. The author simply does not present enough group, ring, and field theory. There is another problem. The course in abstract algebra is often used to introduce the student to standards of mathematical rigor and proof. The style of Childs' book is too informal to serve this purpose well. An ideal set-up would be a two-semester course, the first out of Child's book, and the second out of a more conventional text. Alternatively, it could be used as the basis of a course for nonmath majors, students who would find the material interesting in itself, and who are not in need of the foundations of abstract algebra as the basis for the study of more advanced mathematics.

Finally, the book is ideal for independent reading. It might be especially suitable for someone who is acquainted with theory but has neglected computations. Calculations and examples not only illuminate theory, they suggest theory. Moreover, mastering a new algorithm is often just plain fun. Perhaps that is why Child's book is not only instructive, it is very entertaining.

Calculus and Analytic Geometry. By C. H. Edwards, Jr., and David E. Penney. Prentice-Hall, Englewood Cliffs, New Jersey, 1982. xvi + 895 pp. $29.95.

PETER ROSENTHAL
Department of Mathematics, University of Toronto, Toronto, Ontario M5S 1A1, Canada

It's spring and our thoughts turn to... choosing a calculus text for next year. We sit quietly, staring at the 4000 pounds of textbooks that aggressive book publishers have sent us. What was wrong with the old Courant? Even the old (pre-fourth) edition of Thomas wasn't so bad.

The sales representative of Prindle, Weber & Schmidt walked into my office and said, "I'm glad to see you're using Swokowski's *Calculus* this year; how do you like it?"

"Nothing personal," I answered, "but it stinks."

"Well," the salesman replied, "It's the best-selling calculus book in North America."

Apparently the salesman was correct; *Calculus with Analytic Geometry* by E. Swokowski is among the best-selling calculus books. I think I was correct too: it stinks. Many of us have apparently been allowing our choice of calculus text to be determined more by the hard-sell techniques of the publishers than by pedagogical considerations.

In selecting calculus texts we get involved in discussions of various points: the level of the text,

the quality of the examples, the coverage of various topics, etc. Unfortunately, as the popularity of Swokowski proves, we often pay little attention to the exposition itself. At a glance, Swokowski looks reasonable as a text for a middle-level calculus course. Actual reading of the book, however, reveals it to be riddled with statements and explanations that are cumbersome, misleading, or both. I mention some examples below; when reading them I suggest that you imagine you are a student who doesn't already understand the concept that Swokowski is attempting to explain.

On page 369 (second edition), after the natural logarithm has been defined, we encounter:

(8.9) DEFINITION.

> The natural exponential function, denoted by exp, is defined by
>
> $$\exp x = y \quad \text{if and only if} \quad \ln y = x$$
>
> for all x, where $y > 0$.

The next theorem specifies the relationship between ln and exp.

(8.10) THEOREM.

> The natural logarithmic and natural exponential functions are inverse functions of one another.

Proof. According to Definition (1.23) we must prove

$$\begin{array}{lll} \text{(i)} & \ln(\exp x) = x & \text{for every } x, \text{ and} \\ \text{(ii)} & \exp(\ln y) = y & \text{for every } y > 0. \end{array}$$

These statements follow immediately from Definition (8.9), for if $\exp x = y$, then $\ln y = x$, and substitution of $\exp x$ for y gives us (i). Similarly, if $\ln y = x$, then $\exp x = y$, and substitution of $\ln y$ for x gives us (ii).

First note (did you miss it because you assumed he said what he should have said?) the phrase "for all x, where $y > 0$." How many y's are involved? If he insists on an awkward phrasing, he could at least say something like "for all $y > 0$ and all x." Also, think about the sentence, "The next theorem specifies the relationship between ln and exp." Did he *define* exp as the inverse of ln or not? What kind of an idea is the "proof" of Theorem 8.10 supposed to convey?

Physicists undoubtedly appreciate Swokowski's description of g as "the force of gravitational acceleration" whose magnitude is "32 ft sec^2" (curious units for force), which leads to using Newton's Second Law to get the fact that $r''(t) = g$ if r is the position vector of an object acted upon by gravity; (Swokowski, page 733—wouldn't it be so much simpler if g represented the acceleration itself!)

What do you think of the following (from page 793):

> For certain values of θ the directional derivative may be positive (that is, $f(x,y)$ may increase), or negative ($f(x, y)$ may decrease), or it may be 0.

I wonder, for those values of θ where it is none of the above, what is it?

Or (from page A 9):

> If $f(x)$ has a limit as x approaches a, then that limit is unique.

Are the other limits unique too?

There are many other instances where Swokowski's use of language makes a simple concept appear difficult. Sometimes it's hard to be sure what he's saying even when you know the material thoroughly. What, for example, could he possibly mean by the assertion that "It is also possible to give a strictly algebraic proof" of the fact that limits are unique (page A9)?

This book would not be a good one even if the ambiguities and misleading statements were removed. The proofs are written in a peculiarly formal and singularly uninstructive style. Previous results are often cited by number without attaching a name, even if the name is standard. Thus one has to search through the 1000 pages to find out what is meant by "Applying 16.19," which is code for "By the chain rule."

Some of us justify using an admittedly bad book by saying "The students don't read the book anyway; its only purpose, really, is for assigning problems." Such a claim is not an adequate justification for forcing students to buy huge books that are so financially and physically taxing. If a text is to be merely a source of homework problems, then it should be replaced by a list of problems or, perhaps, Schaum's outline Calculus.

I think that it *is* useful to have a text that can be read by students who want additional explanations of some topics. It is unfortunate, however, that we have allowed the publishers to limit us to choosing big expensive books. If we do choose such a book, it should at least be clear and correct.

There is a new entry in the calculus sweepstakes, *Calculus and Analytic Geometry*, by C. H. Edwards, Jr. and David E. Penney (Prentice-Hall, Engelwood Cliffs, N. J., 1982). It is big and expensive, but it is also very readable and clear, and rarely misleading. It covers the standard topics in the standard order but livens it up with a real variety of applications, most of which are described in language that is simple, careful and also interesting. The explanations of the basic ideas of calculus are also very good, directed towards the student whose interest is in applications of calculus to physical or social science. It really seems possible that a student who is only taking calculus because some bureaucrat decreed that he or she had to will, in spite of everything, find some of the problems intriguing. Students who have a genuine curiosity about engineering, physical science or the more mathematical aspects of social science will certainly find something of interest in this textbook. There is not enough emphasis on the beauty of mathematics as such for this to be a suitable textbook for the most mathematically-inclined students, but it is quite good (infinitely better than Swokowski) for all others. Anyway, for prospective mathematicians Spivak's *Calculus* is close to perfect.

Edwards and Penney is not without faults. In a few places rough statements meant to be intuitive are also misleading. One such that might be mentioned, since it occurs in other calculus books too, is the rough definition of limit that "$F(x)$ gets closer and closer to L as x gets closer and closer to a" (they follow this with a precise definition). Such phrasing suggests that $|F(x) - L|$ is a monotone function of $|x - a|$. It's hard to convince students who think carefully that

$$\lim_{x\to 0} x \sin \frac{1}{x} = 0$$

after they have read such a "rough definition." Why not say "$F(x)$ is very close to L when x is very close to a" (as Edwards and Penney also do), and then go on to the precise definition without introducing a misleading sentence? Also, Edwards and Penney sometimes (e.g., see pages 108 and 504) give a proof of one theorem using a theorem whose proof is given later in the text. While this is logically acceptable, since the theorem proved later does not use the theorem proved with its help, there is serious danger that students will extrapolate incorrect circular methods of reasoning from such instances in a text. Some might not like terminology like "the closed interval maximum-minimum method," especially when it is implied that the terminology is absolutely standard. It would be desirable to have an appendix that includes proofs of theorems such as the existence of integrals of continuous functions; in any class there might be a few students who would be curious about such proofs.

The above criticisms of Edwards and Penney are quite minor. I would wholeheartedly recommend Edwards and Penney if it were smaller and less expensive. Most (perhaps 90%?) calculus students do not continue beyond one year of mathematics, and thus would probably

cover at most 594 of the book's 895 pages. Moreover, they do not require a hardcover expensively-produced textbook that would last for years, especially when second editions are planned in order to destroy the resale value. Also, the cost and bulk of the text are increased by enormous margins.

A modest proposal to Edwards, Penney and Prentice-Hall: produce a two-volume, soft-cover cheap edition. The first volume could contain the first twelve chapters, plus perhaps a brief introduction to partial derivatives. This volume could easily become the best-selling first year calculus book. Then volume two would be available for those students who continue beyond first-year calculus. Neither volume should sell for more than ten dollars.

Introduction to Functional Analysis: Banach Spaces and Differential Calculus. By Leopoldo Nachbin. Marcel Dekker, New York, 1981.

JOE DIESTEL
Department of Mathematics, Kent State University, Kent, OH 33142

The term "functional analysis," as it was understood by the fathers of the area, referred to analyzing function spaces through naturally defined functionals acting on these spaces. By skillfully unearthing invariants whose code could be deciphered by means of these functionals, one could often solve some functional equation the natural domain of which was a judiciously chosen function space. Throughout the first decades of this century, this "soft" approach met with frequent success in such diverse areas as integral equations, minimal surfaces, partial differential equations, harmonic analysis and moment problems.

During the twenties, the spectral theory of operators had stunning applications to problems uniquely susceptible to analysis in Hilbert spaces. The appearance in 1932 of John von Neumann's "Mathematische Grundlagen der Quantenmechanik" and Marshall Stone's "Linear Transformations in Hilbert Spaces and Applications in Analysis" heralded the emergence of operator theory (on Hilbert spaces) as a mature area of mathematical inquiry separate from, but intimately related to, what is now known as linear functional analysis.

At that time, general linear functional analysis experienced its first formative stages of development. Many of the tricks of the "soft analysis" trade crystallized. General principles were formulated. A host of techniques evolved for approaching linear problems more general than those amenable to Hilbert space analysis. Three basic existential principles were soon recognized:

THE HAHN-BANACH EXTENSION THEOREM. *A linear continuous functional on a vector subspace of a normed space admits of a linear continuous extension to the whole space.*

THE BANACH-STEINHAUS THEOREM (alias, THE PRINCIPLE OF UNIFORM BOUNDEDNESS). *Any family of continuous linear operators between Banach spaces that is pointwise bounded on the unit ball is uniformly bounded thereupon.*

THE OPEN-MAPPING THEOREM. *A linear continuous operator from one Banach space onto another maps open sets onto open sets.*

In 1932 the French translation of Stefan Banach's "Opérations Linéaires" appeared. In it, these three theorems were showcased as cornerstones for the structure of functional analysis. After formulating each principle in its most general, palatable version, Banach provided a full cadre of applications of the principle; the effect was to assure their role in the study of linear problems for all time.

Throughout the thirties and forties the boundaries of functional analysis continuously ex-

cover at most 594 of the book's 895 pages. Moreover, they do not require a hardcover expensively-produced textbook that would last for years, especially when second editions are planned in order to destroy the resale value. Also, the cost and bulk of the text are increased by enormous margins.

A modest proposal to Edwards, Penney and Prentice-Hall: produce a two-volume, soft-cover cheap edition. The first volume could contain the first twelve chapters, plus perhaps a brief introduction to partial derivatives. This volume could easily become the best-selling first year calculus book. Then volume two would be available for those students who continue beyond first-year calculus. Neither volume should sell for more than ten dollars.

Introduction to Functional Analysis: Banach Spaces and Differential Calculus. By Leopoldo Nachbin. Marcel Dekker, New York, 1981.

JOE DIESTEL
Department of Mathematics, Kent State University, Kent, OH 33142

The term "functional analysis," as it was understood by the fathers of the area, referred to analyzing function spaces through naturally defined functionals acting on these spaces. By skillfully unearthing invariants whose code could be deciphered by means of these functionals, one could often solve some functional equation the natural domain of which was a judiciously chosen function space. Throughout the first decades of this century, this "soft" approach met with frequent success in such diverse areas as integral equations, minimal surfaces, partial differential equations, harmonic analysis and moment problems.

During the twenties, the spectral theory of operators had stunning applications to problems uniquely susceptible to analysis in Hilbert spaces. The appearance in 1932 of John von Neumann's "Mathematische Grundlagen der Quantenmechanik" and Marshall Stone's "Linear Transformations in Hilbert Spaces and Applications in Analysis" heralded the emergence of operator theory (on Hilbert spaces) as a mature area of mathematical inquiry separate from, but intimately related to, what is now known as linear functional analysis.

At that time, general linear functional analysis experienced its first formative stages of development. Many of the tricks of the "soft analysis" trade crystallized. General principles were formulated. A host of techniques evolved for approaching linear problems more general than those amenable to Hilbert space analysis. Three basic existential principles were soon recognized:

THE HAHN-BANACH EXTENSION THEOREM. *A linear continuous functional on a vector subspace of a normed space admits of a linear continuous extension to the whole space.*

THE BANACH-STEINHAUS THEOREM (alias, THE PRINCIPLE OF UNIFORM BOUNDEDNESS). *Any family of continuous linear operators between Banach spaces that is pointwise bounded on the unit ball is uniformly bounded thereupon.*

THE OPEN-MAPPING THEOREM. *A linear continuous operator from one Banach space onto another maps open sets onto open sets.*

In 1932 the French translation of Stefan Banach's "Opérations Linéaires" appeared. In it, these three theorems were showcased as cornerstones for the structure of functional analysis. After formulating each principle in its most general, palatable version, Banach provided a full cadre of applications of the principle; the effect was to assure their role in the study of linear problems for all time.

Throughout the thirties and forties the boundaries of functional analysis continuously ex-

panded (with a resulting loss of clear definition). Each success resulted in still more adventurous forays into formerly alien territory. Gelfand's investigations into the structure of commutative Banach algebras reunited general linear functional analysis with operator theory to produce, among other things, a startling proof of the spectral theorem for bounded normal operators. Gelfand theory was used to study locally compact abelian groups, a new proof of Pontryagin duality resulted. Fourier analysis on locally compact abelian groups became a viable subject, and abstract harmonic analysis was born.

After the war, the French took up where the Polish school of functional analysts left off, and began a series of intensive investigations into the structure of linear topological spaces, especially spaces of continuously differentiable functions and their duals. Laurent Schwartz developed the theory of distributions (a theory anticipated by others but unquestionably brought to fruition by Schwartz). The stage was set for a highwater mark in the history of functional analysis: the discovery by Bernard Malgrange and Leon Ehrenpreis that all homogeneous linear partial differential equations with constant coefficients have fundamental distributional solutions. Their proof is a tour-de-force in the use of the Hahn-Banach Theorem.

As the sixties dawned, the tools that a young functional analyst needed were diverse—as were the potential applications. Choquet theory wed linear functional analysis with operator theory; it made serious measure theory a valued ally of the functional analyst. Probabilistic techniques and motivations invaded linear analysis and operator theory; complex analysis provided a number of interesting problems that could be recast and solved in a functional analytic framework. Practically every area of analysis provided tractable problems, its own peculiar techniques, and its own intuitions for the functional analyst to test his wares.

These developments paid dividends. Long thought to be unapproachable, the classical problems in Banach spaces were attacked with renewed vigor. (Only a few of the problems posed by Banach in his monograph remain unsolved.) Moreover, serious applications of the structure theory of Banach spaces have been found in harmonic analysis, probability theory, interpolation theory, approximation theory and the distribution of eigenvalues of operators on Hilbert spaces.

The study of operators on a Hilbert space has undergone several dramatic developments. The latest has mated operator theory with K-theory and resulted in serious affairs with differential geometry and algebraic topology. The subject "functional analysis" has come to include a veritable potpourri of mathematical pursuits.

It is the very breadth of the area that allows for a rich variety of introductory books to be written. Many of the most famous functional analysts have contributed important basic texts. Now, Professor Leopoldo Nachbin has added his name to the list.

Professor Nachbin's *Introduction to Functional Analysis*: *Banach Spaces and Differential Calculus* is aimed, remarkably enough, at advanced undergraduate and beginning graduate students from economics, engineering, physics and, perhaps, even mathematics. It aims at acquainting the students with elementary facts about normed linear spaces and the theory of differentiation of functions between such spaces.

Though the three basic principles are stated within the text, no serious undertaking is made to demonstrate their power. Perhaps functional analysis has reached the point where the initial contact with the subject can be effective by just acquainting the student with the main terms and most elementary facts. Perhaps this is so. But, quite frankly, it seems that Professor Nachbin sold the subject (and himself) a bit short. Had he only *stated* the Three Principles and used them at several critical junctures, he might well have written THE classic introduction for the type of audience he was addressing. His writing style is so relaxed and clear he just might have pulled off the missionary accomplishment of twentieth century mathematics: the conversion of a whole horde of economists, engineers, and physicists to functional analysis! By not exhibiting the fundamental principles playing their natural roles in functional analysis, Professor Nachbin has made it more difficult for the serious reader to begin on more earnest study of the subject. With all that one now needs to know, this is unfortunate indeed.

LETTERS TO THE EDITOR

Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

In a recent note, W. M. Snyder [this MONTHLY, 89 (1982) 462–466] considered prime divisors of numbers of the form $(b^n - 1)/(b - 1)$ with integers $n, b > 1$. I wish to point out that the main result of this note can be deduced easily from a classical theorem due to K. Zsigmondy [Monatsh. Math. Phys., 3 (1892) 265–284] and proved again by G. D. Birkhoff and H. S. Vandiver [Ann. of Math., 5 (1904) 173–180]. Furthermore, a crucial result on cyclotomic polynomials in Snyder's note can be proved in a quicker and more elementary fashion.

Snyder's main result says that $(b^n - 1)/(b - 1)$ has a prime divisor $p \equiv 1 \bmod n$ if and only if $n > 2$ or $n = 2$ and $b + 1$ is not a power of 2. According to the theorem of Zsigmondy, for any $n > 2$ and $b > 1$ (except in the case $n = 6$, $b = 2$) there exists a prime divisor p of $b^n - 1$ such that p does not divide any of the integers $b^m - 1$ with m being a proper divisor of n. In particular, p does not divide $b - 1$, hence p divides $(b^n - 1)/(b - 1)$. The properties of p imply also that n is the multiplicative order of $b \bmod p$, and so n divides $p - 1$, i.e., $p \equiv 1 \bmod n$. The remaining cases $n = 2$ and $n = 6$, $b = 2$ in Snyder's result are trivial.

The original proof of Snyder's result uses elementary number theory and known properties of cyclotomic polynomials, except for the proof of the following fact in which extensive use is made of algebraic number theory: *if $\Phi_n(x)$ is the nth cyclotomic polynomial over the rationals and q is a prime dividing $\Phi_n(b)$ but not n, then $q \equiv 1 \bmod n$.* To prove this in an easier way, we note that since q does not divide n, the polynomial $\Phi_n(x)$ considered mod q yields the nth cyclotomic polynomial over the finite field $\mathbb{F}_q$ of integers mod q, whose roots are the primitive nth roots of unity over $\mathbb{F}_q$. Now $\Phi_n(b) \equiv 0 \bmod q$ implies that $\mathbb{F}_q$ contains a primitive nth root of unity, that is, an element of multiplicative order n, and so n divides $q - 1$, i.e., $q \equiv 1 \bmod n$.

Harald Niederreiter
Mathematical Institute, Austrian Academy of Sciences,
Dr. Ignaz-Seipel-Platz 2, A-1010 Vienna, Austria

Editor:

Since the publication of my note "Factoring Repunits" [this MONTHLY, 89 (1982) 462–466], I have received several letters indicating to me that my results are not new. This culminated recently with a "Letter to the Editor" by H. Niederreiter who also showed that the results could be obtained quite easily by elementary methods.

It is very unfortunate that my article was ever published.

To make matters worse, there is an error in the proof of the last lemma. I would like to thank Frank Schmidt from SIU at Carbondale for pointing this out to me. On page 466, the 7th line down "$d \equiv 1$, mod p implies $q \equiv 0 \bmod p - 1$", the conclusion should have been $p - 1 \equiv 0 \bmod q$. Because of this error the lemma, although still correct, needs some changes in the proof. In light of the present circumstances, however, it seems inappropriate for me to bother to correct the proof here. I suggest that anyone interested in the details correct the proof as a fairly easy exercise, or, perhaps, consult the literature for a correct proof.

W. M. Snyder
Department of Mathematics
University of Maine
Orono, Maine 04469

Editor:

In 1976, the Nepali Mathematical Sciences Report came into existence. It is the first mathematics journal to be developed and published in the Kingdom of Nepal. Since 1976, the journal has been published twice a year. Each issue has been about fifty pages long with seven or eight articles ranging over a variety of topics. For the most part these articles have been written by a small group of mathematicians who are located on a campus of the national university just outside Kathmandu. Occasionally, articles written by non-Nepalis from universities in nearby northern India or by mathematicians on assignment from other countries have been published.

Mathematical terminology in the Nepali language is in a developmental stage. Each issue of the journal contains a glossary section, which is devoted to the construction of a consistent set of terms to be used in curriculum development at all levels. To enhance the general problem solving capability of its readers, most issues contain a problem section similar to that found in the MONTHLY. Through these sections the journal serves as a strong educational device to raise the level of mathematical knowledge and understanding throughout the Kingdom.

My purpose in writing to you is two-fold—to inform the readers of the MONTHLY of the existence of this journal and to request their help in providing articles for it. Assistance is needed if the journal is to survive. Expository articles will be very useful. Those that contain a historical perspective are most welcome. I have been one of the editors of the report since its inception and will be glad to receive and forward articles from interested readers. The format should be similar to that used for articles for the MONTHLY.

George F. Feeman
Department of Mathematical Sciences
Oakland University
Rochester, Michigan 48063

MISCELLANEA

113.

Where have all the residents gone? Gone to infinity every one.

According to Gamow (1947/1971, p. 17) and Rucker (1982, p. 73), Hilbert's Hotel has an infinite number of rooms numbered with the integers $1, 2, 3, \ldots,$ each occupied by one resident. Moreover, they can make room for a finite number n of newcomers by each moving "up" n places, or for a countable number of newcomers by each doubling the number of his room. Consider now a slightly different paradox.

The manager, wishing to have the rooms cleaned by an infinite number of chambermaids, at 10 a.m. asks each of the residents to move up one room in one hour, another room in the next half-hour, another room in the next quarter-hour, and so on. By noon all the rooms are empty and the chambermaids begin scrubbing the floors. Where have all the residents gone and how can they get back?

References

1. G. Gamow, One, Two, Three,..., Infinity, Viking Press and Bantam Books, New York, 1947/1971.
2. R. Rucker, Infinity and the Mind, Birkhäuser, Boston, 1982.

—I. J. Good
Statistics Department
VPI & SU
Blacksburg, VA 24061

Editor:

In 1976, the Nepali Mathematical Sciences Report came into existence. It is the first mathematics journal to be developed and published in the Kingdom of Nepal. Since 1976, the journal has been published twice a year. Each issue has been about fifty pages long with seven or eight articles ranging over a variety of topics. For the most part these articles have been written by a small group of mathematicians who are located on a campus of the national university just outside Kathmandu. Occasionally, articles written by non-Nepalis from universities in nearby northern India or by mathematicians on assignment from other countries have been published.

Mathematical terminology in the Nepali language is in a developmental stage. Each issue of the journal contains a glossary section, which is devoted to the construction of a consistent set of terms to be used in curriculum development at all levels. To enhance the general problem solving capability of its readers, most issues contain a problem section similar to that found in the MONTHLY. Through these sections the journal serves as a strong educational device to raise the level of mathematical knowledge and understanding throughout the Kingdom.

My purpose in writing to you is two-fold—to inform the readers of the MONTHLY of the existence of this journal and to request their help in providing articles for it. Assistance is needed if the journal is to survive. Expository articles will be very useful. Those that contain a historical perspective are most welcome. I have been one of the editors of the report since its inception and will be glad to receive and forward articles from interested readers. The format should be similar to that used for articles for the MONTHLY.

George F. Feeman
Department of Mathematical Sciences
Oakland University
Rochester, Michigan 48063

MISCELLANEA

113.

Where have all the residents gone? Gone to infinity every one.

According to Gamow (1947/1971, p. 17) and Rucker (1982, p. 73), Hilbert's Hotel has an infinite number of rooms numbered with the integers $1, 2, 3, \ldots$, each occupied by one resident. Moreover, they can make room for a finite number n of newcomers by each moving "up" n places, or for a countable number of newcomers by each doubling the number of his room. Consider now a slightly different paradox.

The manager, wishing to have the rooms cleaned by an infinite number of chambermaids, at 10 a.m. asks each of the residents to move up one room in one hour, another room in the next half-hour, another room in the next quarter-hour, and so on. By noon all the rooms are empty and the chambermaids begin scrubbing the floors. Where have all the residents gone and how can they get back?

References

1. G. Gamow, One, Two, Three,..., Infinity, Viking Press and Bantam Books, New York, 1947/1971.
2. R. Rucker, Infinity and the Mind, Birkhäuser, Boston, 1982.

—I. J. Good
Statistics Department
VPI & SU
Blacksburg, VA 24061

The Mathematical Association of America
CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Academic Press, Inc.
Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.

THE DOLCIANI MATHEMATICAL EXPOSITION SERIES.

All seven of the books in this widely acclaimed series contain fascinating problems that have ingenious solutions and/or involve intriguing results. These books will challenge the advanced mathematician as well as stimulate the mathematically talented high school student.

The seven volumes available in this series are:

MATHEMATICAL GEMS I, by Ross Honsberger. xi + 176 pages, Hardbound.
List: $16.50. MAA Member: $12.00.

MATHEMATICAL GEMS II, by Ross Honsberger. ix + 182 pages. Hardbound.
List: $16.50. MAA Member: $12.00.

MATHEMATICAL MORSELS, by Ross Honsberger. xii + 249 pages. Hardbound.
List: $21.00. MAA Member: $16.00.

MATHEMATICAL PLUMS, edited by Ross Honsberger. Articles by R.P. Boas, G. D. Chakerian, H. L. Dorwart, D. T. Finkbeiner, Ross Honsberger, K. R. Rebman, and S. K. Stein. ix + 182 pages. Hardbound.
List: $18.00. MAA Member: $13.50.

GREAT MOMENTS IN MATHEMATICS BEFORE 1650, by Howard Eves. xiv + 269 pages. Hardbound.
List: $22.00. MAA Member: $17.00.

MAXIMA AND MINIMA WITHOUT CALCULUS, by Ivan Niven, xv + 323 pages. Hardbound
List: $24.50. MAA Member: $18.50.

GREAT MOMENTS IN MATHEMATICS AFTER 1650, by Howard Eves, xii + 263 pages. Hardbound.
List: $23.50. MAA Member: $17.50.

SPECIAL TWO-VOLUME **GREAT MOMENTS PACKAGE**—"Great Moments... Before and After 1650
List: $39.00. MAA Member: $29.50.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N. W.
Washington, D. C. 20036

WADSWORTH ADVANCED BOOK PROGRAM

ANNOUNCES:

COMPLEX ANALYSIS AND APPLICATIONS, Second Edition

William Derrick, University of Montana

The extensive revision of this undergraduate text for mathematics, engineering, and physics students features an early introduction to complex integration and a wide choice of applications, including optics, jet flows, and wakes as well as the traditional applications. Each selection contains many worked examples. Solutions to the odd-numbered problems are presented in the back of the book. Complimentary copies are available for adoption consideration.

CONTENTS:

1. Analytic Functions
Complex Numbers and their Algebra □ Polar Representation □ Sets in the Complex Plane □ Continuous Functions of a Complex Variable □ Necessary Conditions for Analyticity □ Sufficient Conditions for Analyticity □ The Complex Exponential □ Complex Trigonometric and Hyperbolic Functions □ Complex Logarithms and Power Functions □ Applications in Optics

2. Complex Integration
Line Integrals □ Green's Theorem and its Consequences □ The Cauchy Integral Formula □ Liouville's Theorem and the Maximum Principle □ The Cauchy-Goursat Theorem (Optional)

3. Infinite Series
Taylor Series □ Uniform Convergence of Series □ Laurent Series □ Isolated Singularities □ Analytic Continuation (Optional)

4. Contour Integration
The Residue Theorem □ Evaluation of Definite Real Integrals □ Evaluation of Improper Real Integrals □ Integrals with Poles on the Real Axis □ Integration of Multivalued Functions (Optional) □ The Argument Principle

5. Conformal Mappings
Geometrical Considerations □ Linear Fractional Transformations □ The Symmetry Principle □ Compositions of Elementary Conformal Mappings □ Fluid Flow □ The Schwarz-Christoffel Formula □ Heat Flow and Electrostatics (Optional) □ Wakes in a Fluid Flow (Optional)

6. Boundary Value and Initial Value Problems
Harmonic Functions □ Dirichlet's Problem □ Applications □ Fourier Series □ Fourier Transforms □ Laplace Transforms □ The Inverse Laplace Transform

---

Other texts available for adoption consideration:

INTRODUCTION TO CLASSICAL REAL ANALYSIS
Karl Stromberg, Kansas State University
Advanced undergraduate introduction to analysis.

APPLIED MATHEMATICS
James Cochran, Washington State University
Beginning graduate-level text for mathematicians, engineers, and physicists.

ELEMENTARY MATHEMATICAL ANALYSIS, Second Edition
Colin Clark, University of British Columbia
Undergraduate text for beginning analysis, analysis for secondary teachers.

INTRODUCTION TO MATHEMATICAL LOGIC, Second Edition
Elliott Mendelson, Queens College of CUNY
Advanced introduction to logic.

INTRODUCTION TO LOGIC
Patrick Suppes, Stanford University
Undergraduate text.

Wadsworth International Group
10 Davis Drive
Belmont CA 94002

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 90, Number 9 November 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

Published by the Association at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

DRAWING SEIFERT SURFACES THAT FIBER THE FIGURE-8 KNOT COMPLEMENT IN S^3 OVER S^1

GEORGE K. FRANCIS
Mathematics Department, University of Illinois, Urbana, IL 61801

Few areas of contemporary mathematics can boast of such an abundance of elementary and compelling examples as low dimensional topology. This is the study of curves, surfaces and 3-manifolds. The intense preoccupation of its practitioners with the concrete, tangible arrangement of plastic forms has led them to adopt a vividly pictorial genre of exposition. The process of inventing these apt examples, call it *descriptive topology*, appears to have its own visual logic and graphical rules. These also deserve serious study. As a pedagogical utility, descriptive topology should be straightforward and technically uncomplicated. It should never distract from the mathematics it means to illustrate. Above all, the topological design of its pictures should be memorable, ready to hand at the tip of a pencil or a piece of chalk.

Apt, easy, unambiguous and memorable pictures are what this article is about. In order not to try the reader's patience or offend editorial hospitality, however, I shall avoid vague generalities and focus on a particular example. Technical details on drawing pictures were outlined in [**5**] and a primer for a general method is in preparation [**6**]. Here I shall tell a picture story about visualizing the fibration over the circle of the figure-8 knot complement. This exercise of the imagination consists of filling up the void of space, closed by a point at infinity to form the 3-sphere, S^3, with a continuous succession of surfaces spanning the knot. That is, through each point not on the knot, K, there will pass a unique copy of a compact, connected, bordered, 2-sided surface, F, called a *Seifert surface*, whose boundary is the knot.

The reasons why one should want to do this, and why it should be possible at all, belong to the fascinating biography of this remarkable knot. It is too long to tell about here. May this brief anecdote serve as an invitation to a rewarding treasure hunt in the literature. For a start, you should look at Rolfsen's masterpiece of descriptive topology [**12**] and then proceed to [**9**]. My own interest in it stems from a project to illustrate some ideas in [**3**] and [**20**].

This example is of particular interest to me because it belongs to an area of topology that already has a highly developed and effective graphical shorthand. This shorthand consists of "schematic" diagrams which are highly abstract, and often terse. Recognizable pictures of familiar shapes, arranged in space as dictated by the diagrams, help us see and remember the encoded information. Sometimes, a particularly well-designed picture can lead directly to an insight that is hard to decode from all the algebra generated in the service of precision. What follows is meant to be readable on an undergraduate level not much beyond advanced calculus. John Stillwell's clear and superbly illustrated text [**17**] covers the topological gaps. The more sophisticated reader, whose indulgence I beg for frequent *abus de langage*, is invited to read into my pictures many more ideas from the theory of fibered knots than can be explained at this level. At the end are "solutions" to some of these implied exercises.

G. K. Francis: Here is my story. I still have my first *Bilderbuch*. Printed, bound, and given to me by my grandfather on my second birthday in 1941, it survived the bombs and turmoil of my early years in Europe. My love of pictures, to look at and, later, to draw more for didactic than aesthetic purposes, filled my notebooks with doodles through my school years in South Bend, Indiana. Ky Fan and Yuri Rainich introduced me to topology and geometry at Notre Dame. But this was the era of Bourbaki, and for a decade I erred blindly in higher dimensions, studying differential topology with Raoul Bott. With his blessings I returned to the Midwest and apprenticed myself to the concrete, the particular, the visible topological analysis of Chuck Titus at Michigan. One summer, in Ann Arbor, I assisted Bernard Morin. He lectured on Thom and Mather and I drew his vivid pictures on the blackboard. I labored next in a multiply-connected labyrinth of geometrical disciplines. The corridors of Altgeld Hall, Urbana, are lined with string and plaster models: mute and dusty memories from the days of Klein and Poincaré. Finally, Bill Abikoff arrived and with him Thurston's topological renaissance. Armed with publications and protected by tenure, I headed out on my own quests. Whether I meet giants or windmills only time will tell.

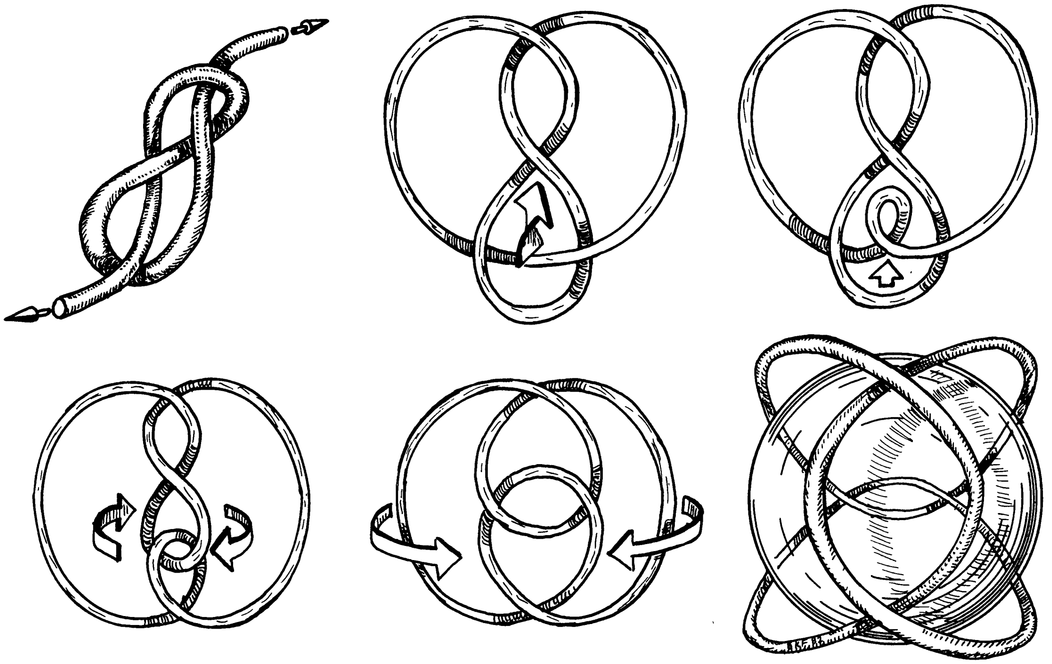


FIG. 1. Views of the figure-8 knot.

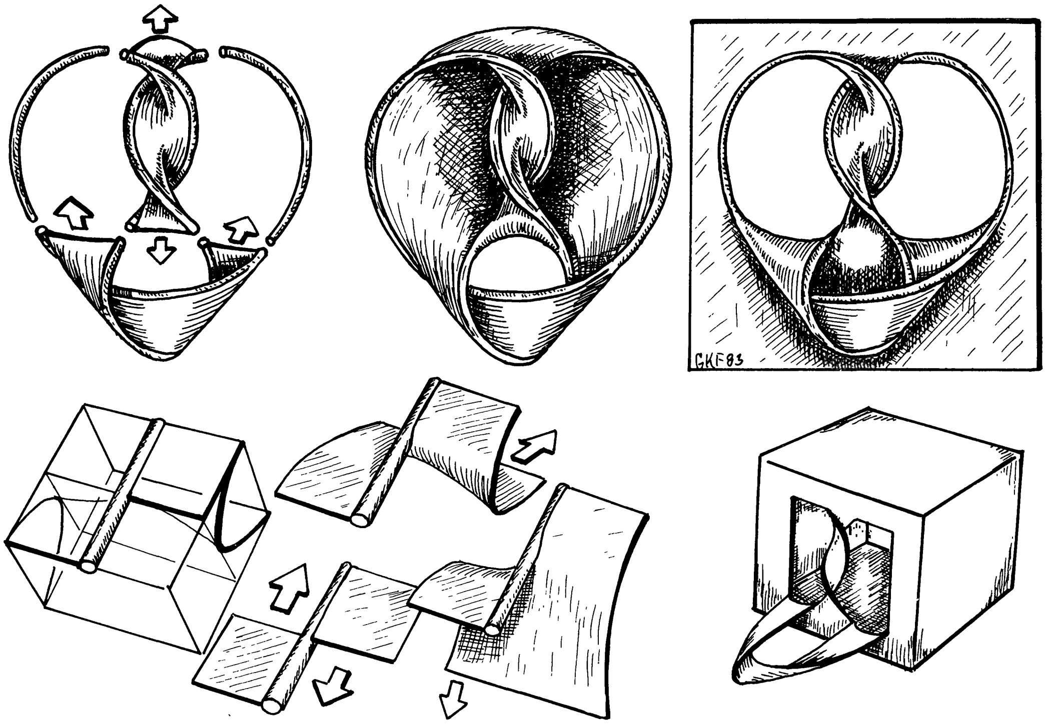


FIG. 2. Seifert surfaces spanning the knot.

Figure 1. Views of the Figure-8 Knot. The figure-eight knot, by sailors "used only to prevent ropes from unreeving; it forms a large knob" [**18**], has been part of "topology" ever since Listing coined this name for the mathematical theory of position [**11**]. Dubbed "four-knot" by Tait, it is the only simple closed knotted curve in space whose picture can be drawn with four, alternating over and under, crossings. Here is a recipe for moving from one of its planar positions 1(12) to the other 1(22). [The abbreviation $n(ij)$ refers to the detail in row i column j of Figure n.] Twist a small loop into a segment of 1(13); this allows you to shrink the large loop 1(21) while untwisting it 1(22). Now swing the large arcs forward, placing the knot in its most symmetric position in space 1(23). A rotation by 90° in the picture plane, followed by a reflection in that plane, moves the knot 1(23) into itself: it is its own "mirror image." Invariance of a geometrical structure under such a turn-reflection (Drehspiegelung) is often useful. Here it will halve our effort in visualizing the fibration of its complement by Seifert surfaces.

Figure 2. Seifert Surfaces Spanning the Knot. It can be quite difficult to see how a surface might span a knot which is drawn in a particular way. For this reason it is wise to experiment with several positions of the knot. Position 1(12) is not the most convenient but serves to illustrate a simple approach to spanning a knot. There is a systematic procedure for doing this, due to Seifert [**14**], based on the combinatorics of the knot projection, but this won't be needed here. With a bit of practice you can learn to "see" twisting ribbons, as well as more or less convex patches, spanning parts of the knot. The difficulty is to see how these patches might fit together. In 2(11) two ribbons, with a full twist each, have come into conflict. They attach to "opposite" sides of the knot. Two ways of resolving this conflict locally are shown in 2(22). 2(21) is a line pattern for this picture. The conflict is resolved globally in 2(12) by continuing the lower ribbon so it fills out a bowl-like patch. This forces the upper ribbon to become a handle or strap across the bowl; the result is Seifert's classic position [**14**]. For a different version, 2(13), continue the upper ribbon of 2(11) to form a plane patch extending to infinity beyond the square border. This forces the lower strap to hang limply from the patch.

The twisted straps of 2(11) could have been merged differently so that the bowl of 2(12) is in front of the upper strap, and the square patch of 2(13) is in front of the lower strap. This specifies four surfaces spanning the knot. These surfaces can be described in terms of 2(23), which is a "cubist" stylization of 2(12). For 2(13) draw the box in front. For the alternate version of 2(12) draw the horizontal strap inside the box; draw the box in front for the alternate of 2(13). A motion, technically called an isotopy in S^3, that takes 2(13) to 2(12) is easy to imagine. Move the infinite, planar patch of 2(13) past infinity. Think of a huge bowl behind the plane of the paper with a square opening into which 2(13) fits. Shrink it to the shallow bowl of 2(12). Now, if you can already imagine an isotopy of 2(12) into a position where the bridge is in back of the bowl, and a similar isotopy through the fourth position back to 2(13), then you have already grasped the gist of this article. Remember, the motion must not involve any tears or self-penetration of the surface, and each surface must differ from every other in all of its interior points. The boundary is, of course, always the knot, K.

Moreover, Seifert's surface turns about K without sliding along it: a page in a ring binder, not a spiral notebook.

Figure 3. Mnemonic Pictures of Fibrations. A *fibration* "over the circle, by Seifert surfaces" of the knot complement $M = S^3 \setminus K$ may be described by a smooth map $f\colon M \to S^1$ of M onto S^1 which is everywhere of maximal rank (equal to 1) and which is "well behaved" where M converges to the knot. 3(11) is a picture mnemonic for this notion. This essentially static structure on S^3 translates to a moving surface, tied to the knot like a sail, by the following bit of differential topology. Orient the circle on which f takes its values and parametrize it by degrees from 0 to 360 for typographical convenience. By the implicit function theorem, the inverse image $F(t) = \{p \in M : f(p) = t\}$ of each $t \in S^1$ is a properly embedded, two-sided, two-dimensional submanifold of M. The surfaces $F(t)$ are oriented by the gradient vector field ∇f, which is everywhere

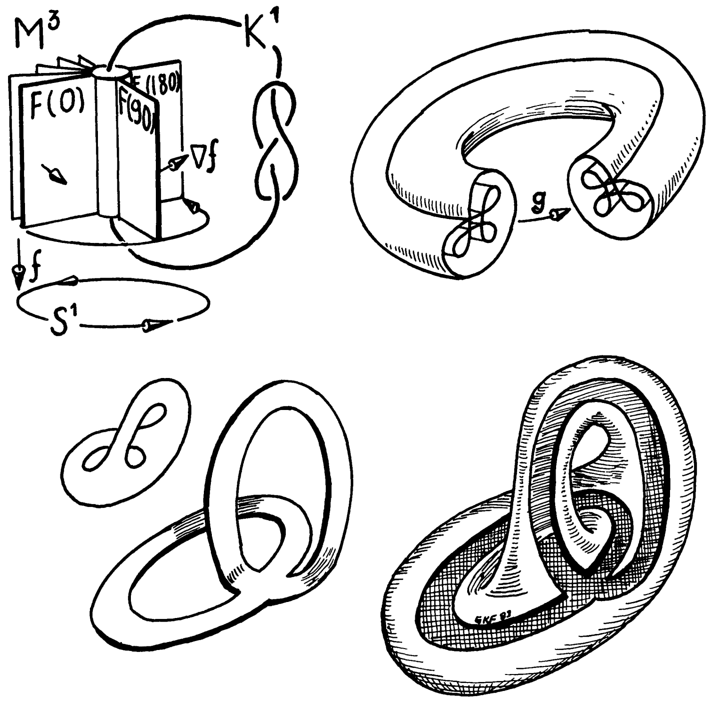


FIG. 3. Mnemonic pictures of fibrations.

perpendicular to them. Choose one of these, $F = F(0)$, as reference surface and integrate along ∇f. This produces a flow on M that moves F through a succession of positions $F(t)$, with the last coinciding with the first. That is, for each $p \in F$ solve the initial value problem

$$dp/dt = \nabla f(p(t)), \qquad p(0) = p.$$

Even though $F(360) = F(0)$, there is no reason to expect the flowlines to close up after one time around; in general $p(360) = p(0)$.

The knottedness of closed flowlines of ∇f is very ably investigated in [3], [4]. The first return map,

$$g: F \to F, g(p(0)) = p(360),$$

called the *monodromy* of the fibration, is a diffeomorphic self-map, or automorphism, of the surface F. This is an example of a Poincaré map for the dynamical system. The manifold may be reassembled from the monodromy by identifying top and bottom of the cylinder $F \times [0, 360]$ by the recipe $(p, 0) \sim (g(p), 360)$. 3(12) is a picture mnemonic for such a mapping torus, $M(g)$, of the automorphism $g: F \to F$.

I have used the conventional symbol, 3(21), for Seifert's surface as cross section of $M(g)$. It represents two annuli joined together. Since such a ribbon neighborhood of two circles crossing once, 3(22), can be joined with a disc, 3(23), to form a whole torus, F is topologically a torus with a disc removed. The difference between 3(22) and all the other versions of Seifert's surface I have drawn is in the twisting of the ribbons. This is a property of the position of the surface in space, not of the surface as an abstract manifold.

Even if the monodromy is not the identity map on F, it is still possible that it is *isotopic* to it. That is, for a smooth map

$$h: F \times [0,1] \to F, \qquad h(p,0) = p, \qquad h(p,1) = g(p),$$

each $h|F \times \{t\}$ is a diffeomorphism. In that case the fibration is said to be trivial because $M(g)$

would be diffeomorphic to $F \times S^1$. As we shall see, this is not the case here. The isotopy class of g is an example of a so called *pseudo-Anosov* mapping in Thurston's classification of the automorphisms of surfaces [**20**].

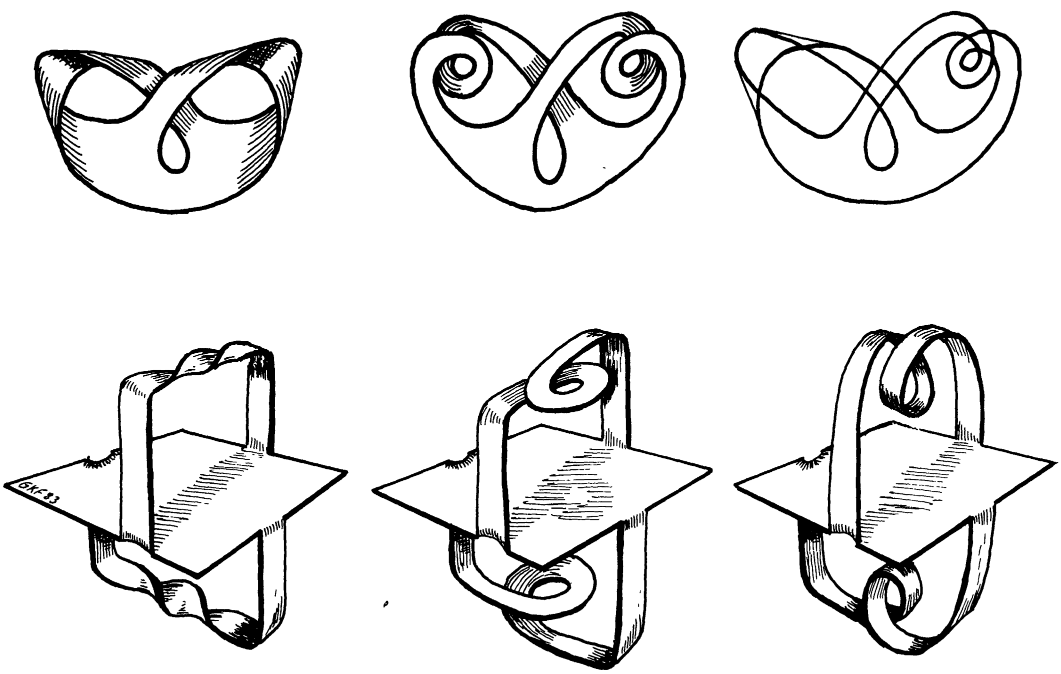


FIG. 4. Alternative views of Seifert's surface.

Figure 4. Alternative Views of Seifert's Surface. Here are some more pictures of F based on its definition as a disc with two oppositely twisted ribbons attached to its circumference. Unexpected equivalences of shapes result when a rubber sheet topologist twists ribbons. The owl 4(12) proceeds from the pussycat 4(11) by a *Seifert move* [**13**] that changes the twisting of ribbons into curling or writhing. [How their line patterns can be remembered from 3(21) is shown in 4(13).] More useful, if less amusing, is the cubist position 4(21), a simpler form of which will concern us for the remainder of the paper.

The importance of such pictorial transmutations in surface topology warrants this brief digression. Curl a strip of paper like the ribbons in 4(23) and pull it taut; it twists like 4(21). A ribbon normal to the curl will writhe, like 4(22), but parallel to a vertical plane. Jim White explains the differential geometry of this phenomenon in [**1**], where its natural manifestation in molecular biology is uncovered. Note how right and left handedness of twist, writhe and curl expresses itself. What remains invariant under an isotopy of a closed ribbon in space is the number of times one edge links the other. The number of obvious twists in the ribbon need not equal this integer; the difference is called its *writhing number*. Both twist and writhe are real numbers associated with the geometry of a particular position in space [**21**]. However, for a picture of a ribbon in general position, Lou Kauffman [**10**] found a direct, visually convenient way of counting twists and curls. The first integer is the net number of times the ribbon twists, ignoring the global overpasses. The second integer is the number of times the ribbon passes over itself from left to right minus the times it does so from right to left, as seen from (either of) its directions. Their sum is the *linking number* of the ribbon edges.

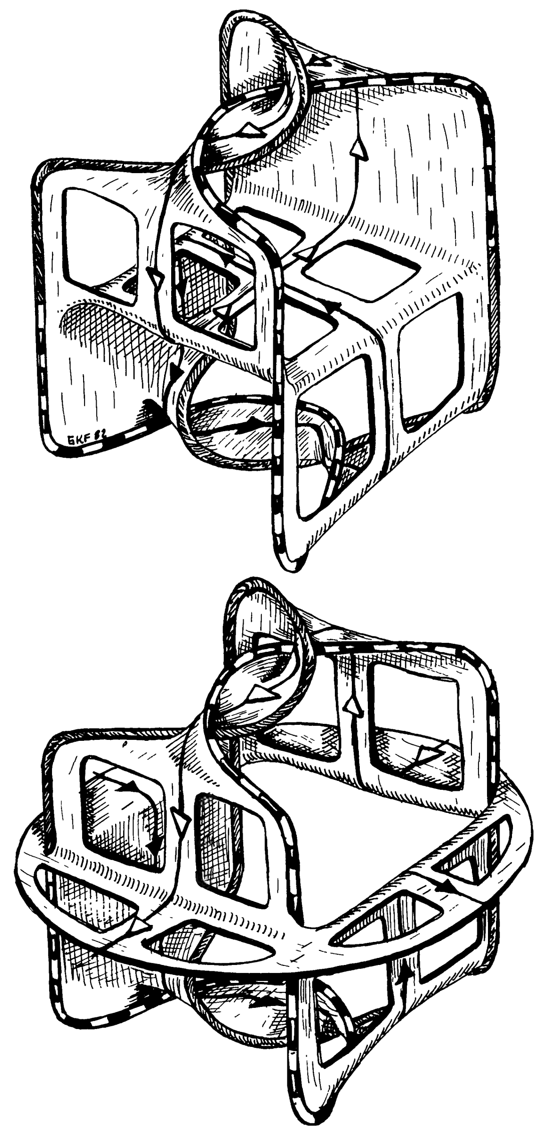


FIG. 5. Two surfaces isotopic in S^3.

Figure 5. Two Surfaces Isotopic in S^3. Of all the positions of K and an initial spanning surface F, 4(21) seems to be the most versatile, once the extraneous bends in K are removed by widening the bridges to form 5(11). To aid the eye, the knot is half striped, half cabled, and 8 windows in F reveal otherwise invisible details. A *window* is a transparent patch on the surface whose projection to the picture plane is a topological disc. The surface is marked by 2 circles, which cross on the horizontal square floor and are directed towards the viewer by white and black arrowheads as they cross the bridges. This is position $F(0)$; $F(180)$ is shown below, 5(21). Here the horizontal floor extends away from the knot on the outside of the box-like frame. The generating circles of $F(180)$ cross at the "center" of this window through infinity. Two simple, closed, directed, once crossing curves on the Seifert surface (torus minus disc) are said to *generate* F because they determine the homeomorphisms of F up to isotopy. In other words, I only need to tell you where the generators go under a mapping in order to identify the isotopy class of the homeomorphism. To visualize the

motion of $F(0)$ to $F(180)$, first imagine the floor of $F(0)$ moving upward, like an elevator, as t advances from 0. As the floor rises past a point on the knot, the surface turns 90° in the vicinity of this point. To see that the surface turns the same way all along the knot, orient the knot so that the striped part heads upward in back and downward in front. You should imagine the surface moving at other places as well, but so slowly that the picture does not change enough to require redrawing it. Soon, however, the floor of $F(80)$ seems to get stuck below the white bridge.

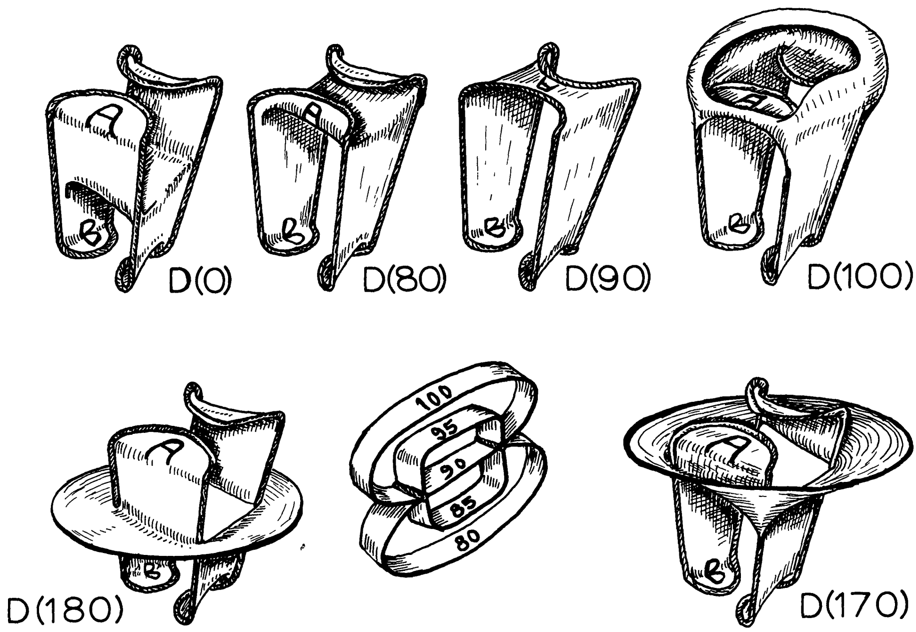


FIG. 6. Relevant positions of the trivial fibration.

Figure 6. Relevant Positions of the Trivial Fibration. Since the motion "through" a bridge will be the hardest part of the visual exercise, let's see what it would look like if the bridges were not there. Fig. 6 shows the relevant stages of a disc D spanning an unknotted curve C bent into a shape similar to the one at hand. At $D(90)$ the spanning disc nearly minimizes area. It bubbles out on top at $D(100)$. Note that during $80 < t < 100$ the surface swings almost 360° about the knot near A, while near B it remains almost stationary. The "almost" is important, because each $D(t)$ must be disjoint from every other. The bubble grows and encloses nearly all of the upper half space at $D(170)$. The balloon bursts at $D(180)$ as the disc passes through infinity. To imagine the rest of the tour, reflect the figures in the horizontal plane, turn them 90° and retrace your steps till $D(360)$ occupies the same position as $D(0)$. This motion describes the fibration over the circle by discs of the complement in S^3 of an unknotted, closed curve C.

Figure 7. The Shear Isotopy. This figure depicts five positions of an annulus A neighboring the white generator of Seifert's surface F as it squeezes through the upper bridge between $t = 80$ and $t = 100$. If the $A(t)$ were superimposed into one drawing, this would correspond to 6(22) in the example just considered. This detail of $F(t)$ was chosen to fit into 2(12) as well as into 5(11). At the bottom is $A(80)$ and $A(100)$ is at the top. The position midway between these, 7(22), is part of $F(90)$. Imagine $A(80)$ shrinking almost into its minimal shape, sliding freely about the two bent rods. To aid the imagination, I have also drawn a position for $t = 85$ on the left, 7(21), and $t = 95$ on the right, 7(23). The shrinking of $A(80)$ to $A(90)$ via $A(85)$ is not quite uniform. The bridge is somewhat stiffer, but catches up quickly between $A(85)$ and $A(90)$.

The annulus $A(80)$ is marked with an equatorial circle, white arrows, and a crosscut, black

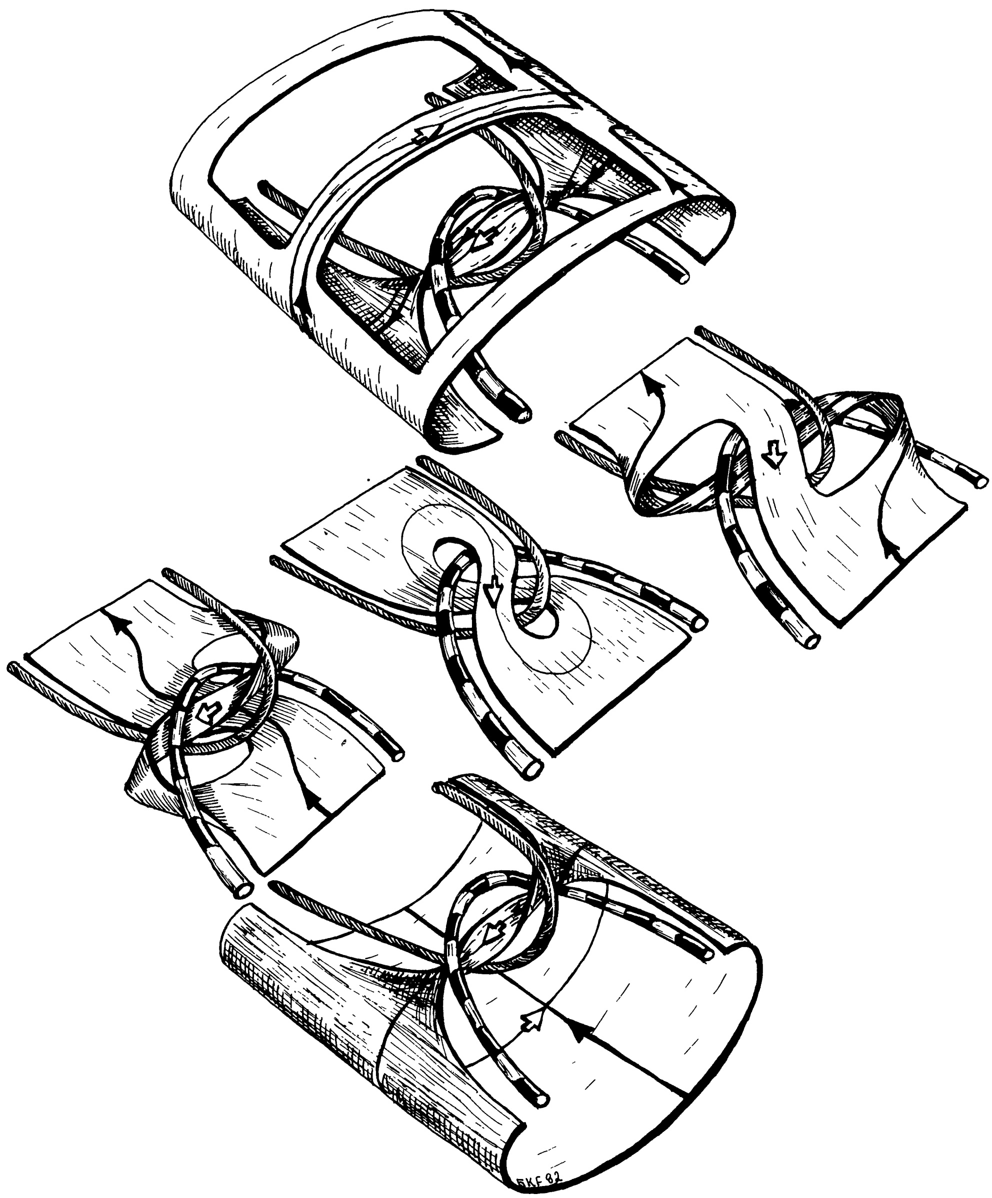


FIG. 7. The shear isotopy.

arrows. On $A(85)$ and $A(95)$ only the crosscut, and on $A(90)$ only the equator has been drawn. You should trace out each of the missing curves to convince yourself that their position on $A(100)$ is as shown. The equator still crosses the bridge towards the front, just as on 5(21). The crosscut, on the other hand, now turns right, following along the equator once around, before crossing it and continuing to the far edge of $A(100)$.

To express the global situation algebraically, let b and w denote the (isotopy classes of the) black and white generators of $F(0)$; B and W those of $F(180)$. Two simple closed curves on F are isotopic if you can slide one on top of the other without leaving the surface. Now draw a hemispherical dome over 5(21) that attaches to the circular rim of the floor. This approximates the

position $F(100)$. As you can see from 7(11), $w = W$ but $b = B + W$. The symbol $B + W$ denotes the class of the closed curve obtained from a pair of generators of F by cutting them at their common point and rejoining them in the obvious way that does not produce conflicting directions on the new curve. This *switching operation* (Umschaltung) at crossings of directed curves and its higher dimensional analogues are one of the fundamental "cut and paste" procedures in topology. It was first used by Gauss in his study of knot diagrams and it is central to Seifert's method for detecting surfaces spanning knots. For the present, however, I only want to switch generators of F.

Thus, the generators (b, w) of $F(0)$ move to $(B + W, W)$ on $F(180)$. This suggests the name *shear isotopy* for the motion that squeezes a surface through a twisted bridge. A better reason for the name, whose full import must regrettably be omitted here, goes as follows. The surface F minus its border curve K may be regarded as a torus with a point removed. The punctured torus may, in turn, be parametrized by the plane $R \times R$ minus the lattice $Z \times Z$ of points with integral coordinates. Points, whose coordinates differ by an element of the translation group $Z \times Z$, name the same point on $F \setminus K$. Up to isotopy, this self map restricted to $F \setminus K$ may, this way, be represented by the linear shear transformation whose matrix is $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

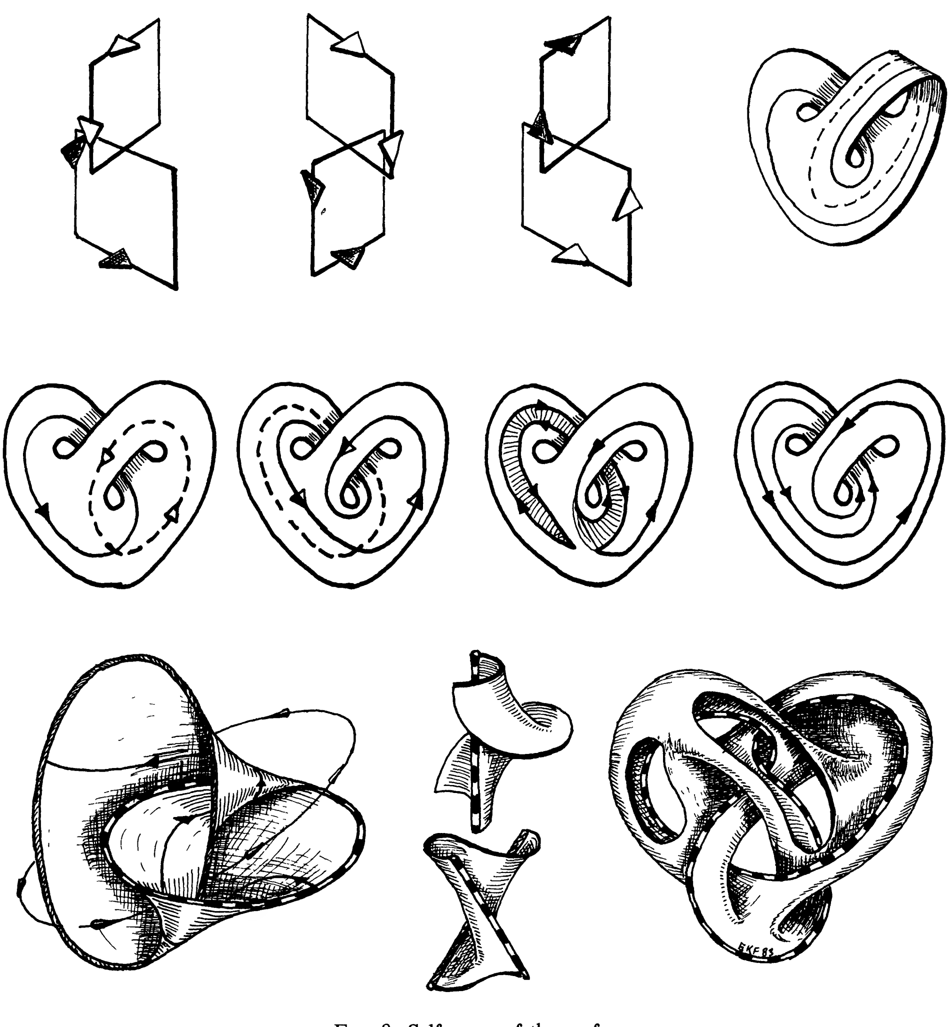


FIG. 8. Self maps of the surface.

Figure 8. Self Maps of the Surface. To compute the fate of $(B + W, W)$ as $F(180)$ moves on to $F(360)$ it is convenient to reverse time and view the motion $F(360, 270, 180)$ through the spectacles of the turn-reflection. To see how this takes (b, w) to $(-w, b)$, turn 8(11) to 8(12) and reflect to obtain 8(13). [The curve $-w$ is w with direction reversed.] A ribbon twisting to the right reflects to one that twists to the left. Therefore, a picture of the shear isotopy $A(280, 270, 260)$ at the lower bridge would look just like $A(80, 90, 100)$ of Fig. 7, but with different markings. Therefore, this conjugation has the effect of taking

$$(b, w) \to (-w, b) \to (-W, B + W) \to (B, W - B).$$

So, since the forward motion from $F(180)$ to $F(360)$ takes $(B, W - B)$ to (b, w), it also takes $W = (W - B) + B$ to $w + b$, and therefore $B + W$ moves to $b + (w + b)$. Since the monodromy map induces an isomorphism of $H_1(F)$, the argument may be summarized algebraically by decomposing its matrix representation thus

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^{-1}\right)^{-1}.$$

These Gauss switches are legitimate because $b + w$ crosses each generator once, up to isotopy on F. This need not be the case on a one-sided surface, such as the Kleinbottle, 8(14), with a disc removed. To see what happens on F, look at the "standard," untwisted version of Seifert's surface, 8(21), with generator b on the left, solid curve, and w on the right, dashed. The solid curve on 8(22) shows why $b + w = w + b$ and 8(23) shows how $(b + w) - b$ is isotopic to w. 8(24) shows the image of b under the monodromy, $g(b) = b + (b + w)$. You should trace into 8(24) a copy of $b + w = g(w)$ that crosses the given curve once, from left to right, showing that the monodromy is orientation preserving.

My picture story is over, but there are a few loose ends to mention. Readers initiated into the beautiful mysteries of Stalling's theory of fibered links, **[15]**, **[16]**, may have noticed that grafting Picture 7, and its mirror image, into Picture 6 is a case of "plumbing" two oppositely oriented "Hopf fibrations," **[7]**, to form Picture 5. I did not recognize this until some time after designing the shear isotopy. The first graft splits the unknot into a pair of circles which link once. If you close up the two bent rods in each detail of Picture 7 into a pair of linked circles, L, the successive positions of the annulus spanning the link, fibers $S^3 \setminus L$ over S^1. If the second graft had also been left handed, the resulting Seifert surfaces, now embedded in S^3 with both straps 2(23) twisted alike, would fiber the complement of the trefoil knot. For a 3-crossing view of this knot, reverse over by under at the lower two crossings in 1(12) and twist the pendant loop, as in 1(21), directly. Hopf's name is more commonly associated with the fibration of S^3 by circles over S^2, induced by the projection of the complex plane to the complex projective line. Though the picture story relating these two notions of fiberings must be told another time, 8(31) suggests how the monodromy is a *Dehn twist*, as defined on p. 158 of **[2]**. One full clockwise turn of the striped rim of this annulus moves the marked crosscut, $c = b - w$, off the twisted bridge. But the monodromy does the same thing, $g(b - w) = (b + w) - w = b$.

For more on this, you should study John Harer's complete account of how every fibered link in S^3, and even in an arbitrary 3-manifold, can be assembled out of such simple pieces. The plumbing of two Hopf links, as drawn in the square saddle shape of 4(21), illustrates a special case of a *Murasugi sum*. This generalized plumbing and its inverse operation preserve several other geometric properties of link complements. Originally treated algebraically, Dave Gabai explores these geometrically in two **[8]** of a series of beautifully illustrated papers on foliations and the topology of 3-manifolds. To him, perceiving how a spanning surface squeezes through unlikely places, as in the shear isotopy, is an exercise every low dimensional topologist should do.

Manipulating knot diagrams is usually a $2\frac{1}{2}$-dimensional procedure; one stays close to the picture plane for reasons of accuracy and convenience in coding the topological information. When truly spatial, rotatable forms are desired, for instance to fix in the imagination what the

diagram represents, some simple, graphical tricks come in handy. How to attach a Seifert surface to the knot is the example at hand. A solid, tubular neighborhood, N, of K carries a canonical structure of simple closed curves on its toroidal surface, T, called a *framing*. The *meridians*, which go around T the short way, span discs inside N. Of all possible curves that go once around T the long way, only the so called *longitudes* span a 2-sided surface in the knot exterior, $X = S^3 \setminus N$. The Meyer-Vietoris isomorphism,

$$H_1(T) \cong H_1(X) \oplus H_1(N),$$

characterizes these curves homologically: $(1,0)$ is the class of meridians, $(0,1)$ is that of the longitudes. Thus Seifert's surface attaches to T along a longitude, and a longitude does not link the knot.

Now from any view, the two contours of T visibly follow once around K the long way, but they need not be longitudes. Spanning surfaces are easier to draw when they are. I used elementary cusp forms, 8(32), to slide 2(12) around the knot until F attaches to T entirely along one contour, 8(33). This led to position 7(22), midway between 7(31) and 7(11) in the shear isotopy.

Acknowledgements. My pictures owe much to the critical eyes of the University of Illinois Geometric Potpourri. Technical assistance came from Wolfgang Haken, Lou Kauffman, Jim White, John Stillwell and Dave Gabai. I am greatly indebted to that master of descriptive topology, Jim Callahan, for invaluable advice in preparing this article. I would like to thank Joan Birman for explaining fibered knots to me originally, and Bob Williams for this irresistible challenge prefacing his own set of pictures in the appendix of [4]:

One of the genuinely nontrivial aspects of 3-dimensional geometry-topology is the visualization of the fibration associated to the complement of a fibered knot in $R^3 \subset S^3$.

References

1. W. Bauer, F. Crick, and J. White, Supercoiled DNA, Sci. Amer., vol. 243, no. 1 (July 1980) 118–133.

2. J. Birman, Braids, Links and Mapping Class Groups, Ann. Math. Studies #82, Princeton University Press, 1975.

3. J. Birman, and R. F. Williams, Knotted periodic orbits in dynamical systems I: Lorenz' equations, Topology, 22 (1983) 47–82.

4. ______, Knotted periodic orbits in dynamical systems II: Knot holders for fibered knots, Columbia University preprint, 1982.

5. G. K. Francis, Drawing surfaces and their deformations: the tobaccopouch eversions of the sphere, Math. Modelling, 1 (1980) 273–281.

6. ______, A Topological Picturebook, in preparation, Springer Verlag.

7. J. Harer, How to construct all fibered knots and links, Topology, 21 (1982) 263–280.

8. D. Gabai, The Murasugi sum is a natural geometric operation, I and II, Harvard U. preprints, 1982 and 1983.

9. McA. C. Gordon, Some aspects of classical knot theory, in Knot Theory, Proceedings Plans-sur-Bex, 1977, Lect. Notes Math. #685, Springer Verlag, 1978.

10. L. Kauffman, Linking, University of Illinois at Chicago preprint, 1981.

11. J. Listing, Vorstudien zur Topologie, Göttinger Studien (1847) 811–875.

12. D. Rolfsen, Knots and Links, Publish or Perish, Berkeley, 1976.

13. H. Seifert, Topologie dreidimensionaler sefaserter Räume, Acta Math., 60 (1933) 147–238.

14. ______, Über das Geschlecht von Knoten, Math. Annalen, 110 (1934) 571–592.

15. J. Stallings, On fibering certain 3-manifolds, in M. Fort, Editor, Topology of 3-Manifolds, Proc. U. Georgia Inst., 1961, Prentice Hall, 1962.

16. ______, Construction of fibered knots and links, in Symposium on algebraic and Geometric Topology, Stanford, 1976, A.M.S. Proc. Symp. Pure Math., 32 (1978) 55–60.

17. J. Stillwell, Classical Topology and Combinatorial Group Theory, Grad. Texts Math., Springer Verlag, 1980.

18. P. G. Tait, Knot, entry in Encycl. Britt., 11th ed., 1911.

19. W. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces, Princeton U. preprint, 1976.

20. ______, Three dimensional manifolds, Kleinian groups, and hyperbolic geometry, Bull. Amer. Math. Soc., 6 (1982) 357–381.

21. J. White, Self-linking and the Gauss integral in higher dimensions, Amer. J. Math., 91 (1969) 693–728.

VERY BASIC LIE THEORY

ROGER HOWE
Department of Mathematics, Yale University, New Haven, CT 06520

Lie theory, the theory of Lie groups, Lie algebras and their applications, is a fundamental part of mathematics. Since World War II it has been the focus of a burgeoning research effort, and is now seen to touch a tremendous spectrum of mathematical areas, including classical, differential, and algebraic geometry, topology, ordinary and partial differential equations, complex analysis (one and several variables), group and ring theory, number theory, and physics, from classical to quantum and relativistic.

It is impossible in a short space to convey the full compass of the subject, but we will cite some examples. An early major success of Lie theory, occurring when the subject was still in its infancy, was to provide a systematic understanding of the relationship between Euclidean geometry and the newer geometries (hyperbolic non-Euclidean or Lobachevskian, Riemann's elliptic geometry, and projective geometry) that had arisen in the 19th century. This led Felix Klein to enunciate his Erlanger Programm [**Kl**] for the systematic understanding of geometry. The principle of Klein's program was that geometry should be understood as the study of quantities left invariant by the action of a group on a space. Another development in which Klein was involved was the Uniformization Theorem [**Be**] for Riemann surfaces. This theorem may be understood as saying that every connected two-manifold is a double coset space of the isometry group of one of the 3 (Euclidean, hyperbolic, elliptic) standard 2-dimensional geometries. (See also the recent article [**F**] in this MONTHLY.) Three-manifolds are much more complex than two manifolds, but the intriguing work of Thurston [**Th**] has gone a long way toward showing that much of their structure can be understood in a way analogous to the 2-dimensional situation in terms of coset spaces of certain Lie groups.

More or less contemporary with the final proof of the Uniformization Theorem was Einstein's [**E**] invention of the special theory of relativity and its instatement of the Lorentz transformation as a basic feature of the kinematics of space-time. Einstein's intuitive treatment of relativity was followed shortly by a more sophisticated treatment by Minkowski [**Mk**] in which Lorentz transformations were shown to constitute a certain Lie group, the isometry group of an indefinite Riemannian metric on $\mathbb{R}^4$. Similarly, shortly after Heisenberg [**Hg**] introduced his famous Commutation Relations in quantum mechanics, which underlie his Uncertainty Principle, Hermann Weyl [**W**] showed they could be interpreted as the structure relations for the Lie algebra of a certain two-step nilpotent Lie group. As the group-theoretical underpinnings of physics became better appreciated, some physicists, perhaps most markedly Wigner [**Wg**], in essence advocated extending Klein's Erlanger Programm to physics. Today, indeed, symmetry principles based on Lie theory are a standard tool and a major source of progress in theoretical physics. Quark theory [**Dy**], in particular, is primarily a (Lie) group-theoretical construct.

These examples could be multiplied many times. The applications of Lie theory are astonishing in their pervasiveness and sometimes in their unexpectedness. The articles of Borel [**Bo2**] and Dyson [**Dy**] mention some. The recent article of Proctor [**Pr**] in this MONTHLY discusses an application to combinatorics. Some points of contact of Lie theory with the undergraduate curriculum are listed in §7.

The article of Proctor also illustrates the need to broaden understanding of Lie theory. Proctor did not feel he could assume knowledge of basic Lie theoretic facts. Though hardly an unknown subject, Lie theory is poorly known in comparison to its importance. Especially since it provides

---

Roger Howe received his Ph.D. from the University of California at Berkeley in 1969. His advisor was Calvin C. Moore. He was at the State University of New York at Stony Brook from 1969 to 1974, and since then has been at Yale University. His main research interests are in group representation theory and harmonic analysis, both pure and applied.

unity of methods and viewpoints in the many subjects to which it relates, its wide dissemination seems worthwhile. Yet it has barely penetrated the undergraduate curriculum, and it is far from universally taught in graduate programs.

Part of the reason for the pedagogy gap is that standard treatments **[A]**, **[Ch]**, **[He]** of the foundations of Lie theory involve substantial prerequisites, including the basic theory of differentiable manifolds, some additional differential geometry, and the theory of covering spaces. This approach tends to put a course in Lie theory, when available, in the second year of graduate study, after specialization has already begun. While a complete discussion of Lie theory does require fairly elaborate preparation, a large portion of its essence is accessible on a much simpler level, appropriate to advanced undergraduate instruction. This paper attempts to present the theory at that level. It presupposes only a knowledge of point set topology and calculus in normed vector spaces. In fact, for the Lie theory proper, only normed vector spaces are necessary. This simplification is achieved by not considering general or abstract Lie groups, but only groups concretely realized as groups of matrices. Since such groups provide the great bulk of significant examples of Lie groups, for many purposes this restriction is unimportant.

The essential phenomenon of Lie theory, to be explicated in the rest of this paper, is that one may associate in a natural way to a Lie group G its Lie algebra $\mathfrak{g}$. The Lie algebra $\mathfrak{g}$ is first of all a vector space and secondly is endowed with a bilinear nonassociative product called the *Lie bracket* or *commutator* and usually denoted $[\,,]$. Amazingly, the group G is almost completely determined by $\mathfrak{g}$ and its Lie bracket. Thus for many purposes one can replace G with $\mathfrak{g}$. Since G is a complicated nonlinear object and $\mathfrak{g}$ is just a vector space, it is usually vastly simpler to work with $\mathfrak{g}$. Otherwise intractable computations may become straightforward linear algebra. This is one source of the power of Lie theory.

The basic object mediating between Lie groups and Lie algebras is the one-parameter group. Just as an abstract group is a coherent system of cyclic groups, a Lie group is a (very) coherent system of one-parameter groups. The purpose of the first two sections, therefore, is to provide some general philosophy about one-parameter groups. Section 1 provides background on homeomorphism groups, and one-parameter groups are defined in a general context in §2. Discussion of Lie groups proper begins in §3. Technically it is independent of §§1 and 2; but these sections will, I hope, give some motivation for reading on. Those who need no motivation or dislike philosophy may go directly to §3. There one-parameter groups of linear transformations are defined and are described by means of the exponential map on matrices. In §4 the exponential map is studied, and the commutator bracket makes its appearance. Section 5 is the heart of the paper. It defines and gives examples of matrix groups, the class of Lie groups considered in this paper. Then it defines Lie algebras, and shows that every matrix group can be associated to a Lie algebra which is related to its group in a close and precise way. The main statement is Theorem 17, and Theorem 19 and Corollary 20 are important complements. Finally §6 ties up some loose ends and §7, as noted, describes some connections of Lie theory with the standard curriculum.

Bibliographical note: The arguments of sections 3, 4 and 5 are very close to those given by von Neumann **[Nn]** in his 1929 paper on Hilbert's 5th problem. A modern development of basic Lie theory which incorporates these results is **[Go]**.

1. Homeomorphism Groups

In this section, we use the standard terminology of general topology, as for example in **[Ke]**.

Let X be a set. Then the collection $\mathrm{Bi}(X)$ of bijections from X to itself is a group with composition of mappings as the group law. Now suppose X is in fact a topological space. Then the set $\mathrm{Hm}(X)$ of homeomorphisms from X to itself is a subgroup of $\mathrm{Bi}(X)$. It seems natural to try to topologize $\mathrm{Hm}(X)$. The topology should of course reflect how $\mathrm{Hm}(X)$ acts on X, so that maps close to the identity move points very little. But a topology on $\mathrm{Hm}(X)$ should also be consistent with the group structure of $\mathrm{Hm}(X)$. More precisely and generally, given a group G, if it is to be made into a topological space in a manner consistent with its group structure, the topology it is

given should satisfy two conditions.

(1.1) (i) The multiplication map $(g_1, g_2) \to g_1 g_2$ from $G \times G$ to G should be continuous.
(ii) The inverse map $g \to g^{-1}$ from G to G should be continuous.

A topology on G satisfying these two compatibility criteria is called a *group topology*. A group endowed with a group topology is called a *topological group*. Some standard treatments of topological groups are **[Hn]** and **[P]**.

In short, then, we would like to make Hm(G) into a topological group. This is not so satisfactorily done for completely general X, but if X is locally compact Hausdorff, there is a nice topology on Hm(X), known as the *compact-open topology*. Before defining it, we make some general observations about group topologies. These will simplify the definition.

Given a group G and an element $g \in G$, define λ_g, *left-translation by* g, and ρ_g, *right-translation by* g, to be the maps

$$\lambda_g : G \to G \qquad \rho_g : G \to G \tag{1.2}$$

given by

$$\lambda_g(g') = gg' \qquad \rho_g(g') = g'g^{-1}.$$

For $U \subseteq G$, set

$$gU = \lambda_g(U) \qquad Ug = \rho_{g^{-1}}(U). \tag{1.3}$$

LEMMA 1. *Let G be a topological group, and $g \in G$.*

(a) *The map $\lambda_g : G \to G$ is a homeomorphism. Similarly $\rho_g : G \to G$ is a homeomorphism.*

(b) *If $U \subseteq G$ is a neighborhood of the identity 1_G of G, then gU and Ug are neighborhoods of g. Similarly if $V \subseteq G$ is a neighborhood of g, then $g^{-1}V$ and Vg^{-1} are neighborhoods of 1_G.*

Proof. One checks from the definition of λ_g that λ is a homeomorphism, i.e.,

$$\lambda_g \circ \lambda_h = \lambda_{gh} \tag{1.4}$$

for $g, h \in G$. It follows directly from the condition (1.1) (i) that λ_g is continuous. Likewise, the map $\lambda_{g^{-1}}$ is also continuous. From (1.4) one concludes that

$$\lambda_{g^{-1}} = (\lambda_g)^{-1}. \tag{1.5}$$

Hence λ_g is continuous with continuous inverse, that is, a homeomorphism. The proof for ρ_g is essentially identical.

Since $\lambda_g(1_G) = g$ and $\lambda_g(U) = gU$ by definition, part b) follows since the homeomorphic image of an open set is open. ■

COROLLARY 2. *A group topology is determined by its system of neighborhoods of the identity.*

Proof. Indeed, a topology on G is determined by the collection of neighborhood systems of each point of G. But according to part (b) of the lemma, for a group topology, the system of neighborhoods around a point $g \in G$ is determined by the system of neighborhoods around 1_G. ■

Let us call a topology on G such that all λ_g and ρ_g are homeomorphisms a *homogeneous topology*. Lemma 1 says group topologies are homogeneous. Evidently Corollary 2 applies to all homogeneous topologies, not only group topologies. Thus an obvious question is what conditions must a neighborhood system at the identity satisfy in order that the associated homogeneous topology be a group topology? This question has a simple answer.

LEMMA 3. *A homogeneous topology on a group G is a group topology if and only if the system of neighborhoods of 1_G satisfies conditions* (a) *and* (b) *below.*

(a) *If U is a neighborhood of 1_G, there is another neighborhood V of 1_G such that $V \subseteq U^{-1}$; where*

$$U^{-1} = \{g^{-1} : g \in U\}. \tag{1.6}$$

(b) *If U is a neighborhood of 1_G, there are other neighborhoods V, W of 1_G such that $VW \subseteq U$, where*

$$VW = \{gh : g \in V, h \in W\}. \tag{1.7}$$

Proof. The conditions (a) and (b) are clearly necessary for a topology to be a group topology, since they are one way of stating that the inverse and multiplication maps are continuous at 1_G. We will check this for condition (b). In order for multiplication to be continuous at $1_G \times 1_G$, given a neighborhood $U \subseteq G$ of 1_G, we must find a neighborhood $U' \subseteq G \times G$ of $1_G \times 1_G$ such that for any point (g, g') of U', the product gg' is in U. But by definition of the product topology on $G \times G$, any neighborhood U' of $1_G \times 1_G$ contains a product $V \times W$, where $V, W \subseteq G$ are both neighborhoods of 1_G. But the image of $V \times W$ under the multiplication map is just the set VW defined in (1.7). So condition (b) amounts to continuity of multiplication at the point $1_G \times 1_G \in G \times G$.

Thus to complete the lemma we need to show that if the multiplication and inverse maps are continuous at the identity, and if the topology on G is homogeneous, then they are continuous everywhere. Let U be a neighborhood of 1_G. Then since λ_g is a homeomorphism, gU is a neighborhood of g, and we may write

$$(gu)^{-1} = u^{-1}g^{-1} = \rho_g(u^{-1}) = \rho_g\left((\lambda_{g^{-1}}(gu))^{-1}\right) \quad u \in U. \tag{1.8}$$

Thus on gU, the inverse map is a composition of $\lambda_{g^{-1}}$, the inverse map on U, and ρ_g. Since $\lambda_{g^{-1}}$ and ρ_g are continuous, and $\lambda_{g^{-1}}$ takes g to 1_G, and the inverse map is continuous at 1_G, we see that the inverse map is continuous at g also. The proof that multiplication is continuous everywhere is analogous and is left as an exercise. ∎

We return to the question of topologizing $\mathrm{Hm}(X)$. Corollary 2 allows us to save work in our definition of the topology on $\mathrm{Hm}(X)$ by only defining neighborhoods of the identity map 1_X on X, and declaring by fiat all left or right translates of these neighborhoods also to be open sets. Lemma 3 tells us what we must check to know our definition yields a group topology.

From now on, we take X to be a locally compact Hausdorff space. Let $C \subseteq X$ be compact, and let $O \supseteq C$ be open. Define

$$U(C, O) = \{h \in \mathrm{Hm}(X) : h(C) \subseteq O, h^{-1}(C) \subseteq O\}. \tag{1.9}$$

If $\{C_i\}$, $1 \leqslant i \leqslant n$, are compact subsets of X, and $\{O_i\}$ are open subsets of X such that $C_i \subseteq O_i$, set

$$U(\{C_i\}\{O_i\}) = \bigcap_{i=1}^{n} U(C_i, O_i). \tag{1.10}$$

DEFINITION. Let X be a locally compact Hausdorff space. The *compact-open* topology on $\mathrm{Hm}(X)$ is the homogeneous topology such that a base for the neighborhoods of 1_X consists of the sets $U(\{C_i\}, \{O_i\})$ of equation (1.10).

PROPOSITION 4. *The compact-open topology on* $\mathrm{Hm}(X)$ *is a Hausdorff group topology.*

Proof. Since we have decreed the compact open topology to be homogeneous, we need only check the conditions of Lemma 3 to show it is a group topology. Condition (a) is automatic since the sets $U(C, O)$ are defined to be invariant under the inverse map on $\mathrm{Hm}(X)$. Let us check condition (b). If U_i, V_i and W_i are neighborhoods of 1_X such that $V_iW_i \subseteq U_i$, then evidently

$$\left(\bigcap_i V_i\right)\left(\bigcap_i W_i\right) \subseteq \bigcap_i U_i.$$

Hence since the sets (1.10) are intersections of the sets $U(C, O)$ of (1.9), it will be enough to check condition (b) with the neighborhood U of the form $U = U(C, O)$. Since X is locally compact Hausdorff, we can by a standard separation theorem (cf. [**Ke**, Chap. 5, Theorem 18]) find an open

$O' \subseteq X$ such that the closure C' of O' is compact, and

$$C \subseteq O' \subseteq C' \subseteq O.$$

Then set $V = W = U(C', O) \cap U(C, O')$. If $h_1, h_2 \in V$, we find

$$h_1 \circ h_2(C) = h_1(h_2(C)) \subseteq h_1(O') \subseteq h_1(C') \subseteq O$$

and similarly for $(h_1 \circ h_2)^{-1} = h_2^{-1} \circ h_1^{-1}$. Thus $h_1 \circ h_2 \in U(C, O) = U$, or $VW \subseteq U$ as was to be shown, and the compact-open topology is a group topology on $\mathrm{Hm}(X)$.

To show that a group topology is Hausdorff is a fairly simple matter. We record the relevant observation as a separate result.

LEMMA 5. *Let G be a topological group. Let $H \subseteq G$ be the intersection of all neighborhoods of 1_G. Then H is a normal subgroup of G. Further, G is Hausdorff if and only if $H = \{1_G\}$.*

Proof. Suppose $h_1, h_2 \in H$. Given a neighborhood U of 1_G, we can find neighborhoods V, W of 1_G such that $VW \subseteq U$. Since $h_1 \in V$ and $h_2 \in W$, we see that $h_1 h_2 \in U$. Hence $h_1 h_2 \in H$ also. In similar fashion, one sees that $h_1^{-1} \in H$. Hence H is a group. Since the conjugate gUg^{-1} of a neighborhood of 1_G is again a neighborhood of 1_G, we see that H is also normal in G.

Suppose $H = 1_G$. Then given $g \in G$, we can find a neighborhood U of 1_G such that $g \notin U$. Let V, W be neighborhoods of 1_G such that $VW \subseteq U$. Then gV^{-1} and W are neighborhoods of g and of 1_G, respectively, and are disjoint. Now consider any two points $g_1, g_2 \in G$. Set $g = g_1^{-1} g_2$, and apply the argument above. Translating on the left by g_1, we find $g_1 W$ and $g_2 V^{-1}$ are disjoint neighborhoods of g_1 and g_2, respectively. Hence G is Hausdorff. ■

From Lemma 5 we see Proposition 4 will be proved if we produce for each $h \neq 1_X$ in $\mathrm{Hm}(X)$ a compact C and open O such that $h \notin U(C, O)$. Choose $x \in X$ such that $h(x) \neq x$. Then evidently $h \notin U(\{x\}, X - \{h(x)\})$. ■

REMARK. In fact the compact-open topology on $\mathrm{Hm}(X)$ is better than Proposition 4 indicates. It is complete with respect to an appropriate uniform structure ([**Ke**, Chap. 6]). Also, if X is second countable (hence metrizable), then $\mathrm{Hm}(X)$ is also second countable and metrizable.

2. One-Parameter Groups: Flows and Differential Equations

The real number system $\mathbb{R}$ equipped with addition and its familiar topology is, as the reader may easily check, a topological group.

DEFINITION. A *one-parameter group* of homeomorphisms of (the locally compact Hausdorff space) X is a continuous homomorphism

$$\varphi : \mathbb{R} \to \mathrm{Hm}(X). \tag{2.1}$$

It will be convenient to denote the image under φ of t by φ_t rather than $\varphi(t)$. Thus $\{\varphi_t\}$ is a family of homeomorphisms of X satisfying the rule

$$\varphi_t \circ \varphi_s = \varphi_{t+s} \qquad t, s \in \mathbb{R}. \tag{2.2}$$

Since for each t the map φ_t acts on X, a one-parameter group of homeomorphisms of X is also called an $\mathbb{R}$-action on X, or an action by $\mathbb{R}$ on X.

Given a one-parameter group φ_t of homeomorphisms of X, we can define a map

$$\begin{aligned} \Phi &: \mathbb{R} \times X \to X, \\ \Phi(t, x) &= \varphi_t(x). \end{aligned} \tag{2.3}$$

The fact that $t \to \varphi_t$ is a homomorphism is captured by the identities

$$\begin{aligned} &\text{(i) } \Phi(0, x) = x, \\ &\text{(ii) } \Phi(s, \Phi(t, x)) = \Phi(s + t, x). \end{aligned} \tag{2.4}$$

The continuity of φ is reflected in the continuity of Φ. We state this fact formally. It will perhaps also shed some light on the significance of the compact-open topology on $\mathrm{Hm}(X)$.

LEMMA 6. *Let $\Phi : \mathbb{R} \times X \to X$ be a map. For $t \in \mathbb{R}$, define $\varphi_t : X \to X$ by formula* (2.3) (ii). *Then $\{\varphi_t\}$ is a one-parameter group of homeomorphisms if and only if*

(a) *Φ satisfies identities* (2.4) *and*

(b) *Φ is continuous.*

REMARK. According to this lemma, if our goal were simply to define a one parameter group of homeomorphisms in the quickest way, we could short-circuit the whole discussion of §1 and simply define a one-parameter group of homeomorphisms as a map Φ satisfying the conditions of the lemma. However, that approach seemed unduly formalistic.

Proof. It is a straightforward computation to verify that the identities (2.4) guarantee that for each t the map φ_t is in $\mathrm{Bi}(X)$ and $t \to \varphi_t$ is a homomorphism. Also it is obvious that the maps φ_t will be in $\mathrm{Hm}(X)$ if and only if Φ is continuous in x for each fixed t. Thus the main thrust of the lemma is that $t \to \varphi_t$ is continuous from $\mathbb{R}$ to $\mathrm{Hm}(X)$ if and only if Φ is jointly continuous in t and x. Let us verify this.

Suppose Φ is continuous. Let $C \subseteq X$ be compact, and $O \subseteq X$ be open, with $C \subseteq O$. Choose any $x \in C$. By identity (2.4)(i), the point $(0, x) \in \mathbb{R} \times X$ is in $\Phi^{-1}(0)$. By continuity of Φ a neighborhood of $(0, x)$ is contained in $\Phi^{-1}(0)$. This means there is a neighborhood N of x in X, and $\delta > 0$, depending on x, N, and O, such that $\Phi(t, y) \in 0$ for $y \in N$ and $|t| < \delta$. In other words $\varphi_t(y) \in O$ for $|t| < \delta$ and $y \in N$. Since C is compact, we can find a finite number of $x_i \in C$ such that the associated neighborhoods N_i cover C. Suppose then that $\Phi(t, y_i) \in O$ for $y_i \in N_i$ and $|t| < \delta_i$. Set $\delta = \min \delta_i$. Then we have $\Phi(t, c) \in O$ for all $c \in C$ and $|t| < \delta$. In other words, $\varphi_t \in U(C, O)$ for $|t| < \delta$. Clearly, by repeating this argument for any finite collection of compact C_i's and open O_i's containing them, we can show that $\varphi_t \in U(\{C_i\}, \{O_i\})$ for all sufficiently small t. This shows that $t \to \varphi_t$ is continuous at the origin in $\mathbb{R}$. But now we appeal to the following lemma.

LEMMA 7. *Let $\varphi : G \to H$ be a homomorphism between topological groups. Then φ is continuous if and only if φ is continuous at 1_G.*

The proof of this lemma is left as an exercise to the reader, who will recognize in it the same spirit that informs Lemmas 2, 3, and 5.

To finish Lemma 6, we must show that the continuity of $t \to \varphi_t$ implies continuity of Φ. Choose $(t, x) \in \mathbb{R} \times X$, and set $y = \Phi(t, x)$. Let V be a neighborhood of y. Since φ_t is continuous, we can find a neighborhood W of x, with compact closure $\overline{W}$, such that $\varphi_t(\overline{W}) \subseteq V$. Since φ_t is continuous in t, we can find $\varepsilon > 0$ so that $\varphi_s \in U(\varphi_t(\overline{W}), V)$ for $|s| < \varepsilon$. But then if $(t', w) \in (t - \varepsilon, t + \varepsilon) \times W$, we have

$$\Phi(t', w) = \Phi\big(t' - t, \Phi(t, w)\big) = \varphi_{t'-t}(\Phi(t, w)) \in \varphi_{t'-t}\big(\varphi_t(W)\big) \subseteq V.$$

In other words $(t - \varepsilon, t + \varepsilon) \times W \subseteq \Phi^{-1}(V)$. Since V was an arbitrary neighborhood of y, we see Φ is continuous at (t, x). Since (t, x) is arbitrary, we see Φ is continuous. ■

Consider a one-parameter group φ_t of homeomorphisms of X and the associated map Φ defined by formula (2.3). The map Φ is a function of two variables, t and x, and the maps φ_t are obtained from Φ by temporarily fixing t and letting x vary. If on the other hand we fix x and let t vary, we get a map $t \to \Phi(t, x) = \varphi_t(x)$ which defines a continuous curve in X, traced by the moving point $\varphi_t(x)$. Thus as t varies, each point of x moves continuously inside X, and various points move in a coherent fashion, so that we can form a mental picture of them flowing through X, each point along its individual path. For this reason, a one-parameter group of homeomorphisms of X is also sometimes called a *flow* on X.

The notion of a flow is closely related to the theory of differential equations. Indeed, let $X = \mathbb{R}^n$, and write

$$x = (x_1,\dots,x_n) \qquad x \in \mathbb{R}^n,\ x_i \in \mathbb{R}.$$

Then

$$\Phi(t,x) = \Phi(t,x_1,x_2,\dots,x_n) = (\Phi_1(t,x_1,\dots,x_n),\dots,\Phi_n(t,x_1,\dots,x_n))$$

is a function from $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$. Suppose Φ is not merely continuous, but differentiable. Define

$$f\colon \mathbb{R}^n \to \mathbb{R}^n$$

by

$$f(x) = \left.\frac{\partial \Phi}{\partial t}(t,x)\right|_{t=0}. \tag{2.5}$$

If we differentiate (2.4) (ii) with respect to s, and set $s = 0$, we obtain

$$\frac{\partial \Phi(t,x)}{\partial t} = f(\Phi(t,x)). \tag{2.6}$$

In other words, for fixed x, the map $y_x(t) = \Phi(t,x)$ is a solution of the system of differential equations

$$\frac{dy}{dt} = f(y), \tag{2.7}$$

or

$$\frac{dy_i}{dt} = f_i(y_1,\dots,y_n), \quad \text{for } 1 \leqslant i \leqslant n.$$

The solution y_x of (2.7) is the solution of (2.7) with initial condition $y_x(0) = x$.

The system (2.7) may be pictured geometrically as follows. At each point $y \in \mathbb{R}^n$, one draws the vector $f(y) = (f_1(y), f_2(y),\dots,f_n(y))$. This gives a family of vectors which vary smoothly as y varies; such a family is called a *vector field*. A solution of the system (2.7) is a parametrized curve $c(t)$ in $\mathbb{R}^n$, such that at each point $c(t)$ of the curve the tangent vector $c'(t)$ is the pre-assigned vector $f(c(t))$. The 2-dimensional system

$$\frac{d(x,y)}{dt} = (-y,x)$$

whose solutions are the circles

$$(x(t),y(t)) = (a\cos(\theta_0 + t), a\sin(\theta_0 + t))$$

is illustrated in Fig. 1.

Suppose on the other hand that for each x we have a solution $y_x(t)$ of the system (2.7) with initial condition $y_x(0) = x$. For $s \in R$, consider the function

$$y_{x,s}(t) = y_x(t+s).$$

Differentiation of $y_{x,s}$ shows it also is a solution of the system (2.7), evidently with initial value $y_{x,s}(0) = y_x(s)$. The uniqueness part of the Existence and Uniqueness Theorem for ordinary differential equations **[L]**, **[HS]**, **[R]**, therefore implies that

$$y_x(s+t) = y_{x,s}(t) = y_{y_x(s)}(t). \tag{2.8}$$

If we then set

$$\Phi(t,x) = y_x(t),$$

we find that identity (2.8) translates into identity (2.4) (ii). Of course, the initial condition $y_x(0) = x$ is just identity (2.4) (i). It follows that $\varphi_t(x) = y_x(t)$ defines a one-parameter group of

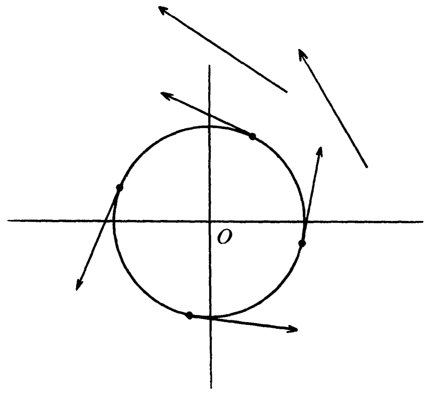


FIG. 1.

homeomorphisms of $\mathbf{R}^n$. For the system of Fig. 1, the map φ_t is just rotation through an angle of t radians.

In summary, we have seen that a (smooth) one-parameter group of diffeomorphisms of $\mathbf{R}^n$ yields solutions of a system of differential equations of the form (2.7), and, conversely, a solution (for all time and all x) of system (2.7) yields a one-parameter group. The two constructs, solutions of systems of ordinary differential equations, and one-parameter groups, thus provide two different points of view on the same mathematical phenomenon. In other words, the notion of one-parameter group provides a geometric and global way of looking at the solutions of a system of ordinary differential equations. As such, it suggests ways of attacking and obtaining information about ordinary differential equations, and it provides a link between systems of ordinary differential equations and more complex geometric objects such as the Lie groups and Lie algebras discussed in the following sections.

3. One-Parameter Groups of Linear Transformations

In this section, we show how one-parameter groups of linear transformations of a vector space can be described using the exponential map on matrices.

Let V be a finite dimensional real vector space. Let $\mathrm{End}(V)$ denote the algebra of linear maps from V to itself, and let $\mathrm{GL}(V)$ denote the group of invertible linear maps from V to itself. The usual name for $\mathrm{GL}(V)$ is the *general linear group* of V. If $V = \mathbf{R}^n$, then $\mathrm{End}(V) = M_n(\mathbf{R})$, the $n \times n$ matrices, and $\mathrm{GL}(V) = \mathrm{GL}_n(\mathbf{R})$, the matrices with nonvanishing determinants.

Let $\| \ \|$ be a norm on V (c.f. [L], [N]). In the usual way there is induced an operator norm, also denoted $\| \ \|$, on $\mathrm{End}(V)$. We recall the definition:

$$(3.1) \qquad \|A\| = \sup\left\{ \frac{\|Av\|}{\|v\|} : v \in V - \{0\} \right\} \quad A \in \mathrm{End}(V).$$

The norm on $\mathrm{End}V$ makes $\mathrm{End}(V)$ into a metric space. Since the determinant is a continuous function on $\mathrm{End}(V)$, we know that $\mathrm{GL}(V)$ is an open subset of $\mathrm{End}V$ (see also (3.6) below), so it also is a metric space.

DEFINITION. A *one-parameter group of linear transformations* of V is a continuous homomorphism

$$(3.2) \qquad M : \mathbf{R} \to \mathrm{GL}(V).$$

Thus $M(t)$ is a collection of linear maps such that

(i) $M(0) = 1_V$, the identity of V,

(ii) $M(s)M(t) = M(s+t) \qquad s, t \in \mathbf{R}$,

(iii) $M(t)$ depends continuously on t.

REMARKS. (a) The topology on $GL(V)$ is easily verified to be a group topology as defined in 1. Thus, for $A \in \operatorname{End} V$ and $r > 0$, set

$$\mathscr{B}_r(A) = \{A' \in \operatorname{End} V : \|A' - A\| < r\}. \tag{3.3}$$

First, the basic formula [**N**, p. 76],

$$\|AB\| \leqslant \|A\|\,\|B\| \tag{3.4}$$

implies that left and right multiplication are continuous. Hence the topology is homogeneous. Then the Neumann formula [**N**, p. 177],

$$(1_V - A)^{-1} = \sum_{n=0}^{\infty} A^n \tag{3.5}$$

valid for A with $\|A\| < 1$ shows that

$$\mathscr{B}_r(1_V))^{-1} \subseteq \mathscr{B}_s(1_V) \tag{3.6}$$

with $s = r/(1-r)$. Similarly the formula

$$(1_V + A)(1_V + B) = 1_V + A + B + AB \tag{3.7}$$

shows

$$\mathscr{B}_r(1_V)\mathscr{B}_s(1_V) \subseteq \mathscr{B}_{r+s+rs}(1_V).$$

Thus all the conditions of Lemma 3 are checked, and we have a group topology.

(b) Furthermore, it is not difficult to verify that the topology defined by the norm coincides with the compact-open topology defined in §1 on GL(V) as a subgroup of Hm(V). This is left as an exercise. Hence this definition of one-parameter group is a special case of the definition of §2.

For $A \in \operatorname{End} V$, define

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!}. \tag{3.8}$$

Since $\|A^n\| \leqslant \|A\|^n$, we see, by the standard estimates in the exponential series, that the series defining $\exp A$ converges absolutely for all A and uniformly on any $\mathscr{B}_r(0)$. Hence exp defines a smooth, in fact analytic, map from End(V) to itself. We will see shortly that in fact $\exp A \in$ GL(V).

PROPOSITION 8. *If A and B in* End V *commute with each other, then*

$$\exp(A + B) = \exp A \exp B. \tag{3.9}$$

Proof. Computing formally we have

$$\begin{aligned}\exp A \exp B &= \left(\sum_{n=0}^{\infty} \frac{A^n}{n!}\right)\left(\sum_{m=0}^{\infty} \frac{B^m}{m!}\right) = \sum_{n,m=0}^{\infty} \frac{A^n B^m}{n!m!}\\ &= \sum_{l=0}^{\infty} \frac{1}{l!}\left(\sum_{m+n=l} \frac{l!}{m!n!} A^n B^m\right)\\ &= \sum_{l=0}^{\infty} \frac{1}{l!}\left(\sum_{k=0}^{l} \binom{l}{k} A^k B^{l-k}\right).\end{aligned}$$

If A and B commute, the familiar binomial formula applies and says

$$(A+B)^l = \sum_{k=0}^{l} \binom{l}{k} A^k B^{l-k}.$$

Substituting this in our formula for $\exp A \exp B$, and noting that all manipulations are valid because the series converge absolutely, we see the proposition follows. ■

COROLLARY 9. *For any $A \in \operatorname{End} V$, the map $t \to \exp(tA)$ is a one-parameter group of linear transformations on V. In particular $\exp A \in \mathrm{GL}(V)$ and $(\exp(A))^{-1} = \exp(-A)$.*

Proof. Since for any real numbers s and t the matrices sA and tA commute with one another, this corollary follows immediately from Proposition 8. ■

The main result of this section is the converse of Corollary 9.

THEOREM 10. *Every one-parameter group M of linear transformations of V has the form*

$$M(t) = \exp(tA) \tag{3.10}$$

for some $A \in \operatorname{End} V$.

The transformation A is called the *infinitesimal generator* of the group $t \to \exp(tA)$. The flow illustrated in Figure 1 is in fact given by a one-parameter group with infinitesimal generator $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

REMARK. Since for v $\in$ V we have

$$(\exp tA)(v) = v + tA(v) + \sum_{n=2}^{\infty} \frac{t^n A^n(v)}{n!},$$

the infinitesimal generator of the one-parameter group $M(t) = \exp(tA)$ can be computed by the formula

$$A(v) = \lim_{t \to 0} \frac{M(t)(v) - v}{t} = \frac{d}{dt}(M(t)(v))\Big|_{t=0}. \tag{3.11}$$

Thus the one-parameter group $M(t)$ is associated by the discussion at the end of §2 to the system of differential equations

$$\frac{dv}{dt} = A(v). \tag{3.12}$$

These equations are of course basic in the theory of linear systems, which is applied in electrical engineering, economics, etc. If we know that $M(t)(v)$ is differentiable, then the existence and uniqueness theorem for differential equations implies Theorem 10, but we do not know *a priori* that $M(t)$ is differentiable. The burden of the proof of Theorem 3 is to get around this ignorance, thereby establishing that a merely continuous map $t \to M(t)$ satisfying the group law (3.2) (ii) is in fact analytic. This is a recurrent theme in Lie theory, and is also expressed in the main theorem (Theorem 17) of this paper. It found its ultimate expression in Hilbert's 5th Problem: to show that if a topological group is locally (i.e., a neighborhood of every point is) homeomorphic to Euclidean space, then the group is in fact an analytic manifold with analytic group law (a Lie group). This problem was resolved positively in the early 1950's by A. Gleason [**G**]. See also [**Ka**], [**MZ**].

We take up now the proof of Theorem 10. It will require some preliminary results.

Let $\mathscr{B}_r(A)$ be the open ball of radius r around A, as defined in formula (3.3).

PROPOSITION 11. *For sufficiently small $r > 0$, the map* exp *takes $\mathscr{B}_r(0)$ bijectively onto an open neighborhood of 1_V in $\mathrm{GL}(V)$. One has $\exp(\mathscr{B}_r(0)) \subseteq \mathscr{B}_s(1_V)$ where $s = e^r - 1$.*

Proof. Let $D\exp_A$ be the differential of exp at A. It is a linear map from $\operatorname{End}(V)$ to $\operatorname{End}(V)$ defined by

$$D\exp_A(B) = \lim_{t \to 0} \frac{\exp(A + tB) - \exp A}{t}.$$

From the definition (3.8) of exp, it is easy to compute that

$$D\exp_0(B) = B.$$

That is $D\exp_0$ is the identity map on $\mathrm{End}(V)$. In particular $D\exp_0$ is invertible. Therefore the first statement of the proposition follows from the Inverse Function Theorem **[L]**, **[R]**. The inclusion $\exp(\mathscr{B}_r(0)) \subseteq \mathscr{B}_s(1_V)$ follows from the obvious termwise estimation of $\exp(A) - 1_V$. ■

REMARK. If one defines

$$\log(1_V - A) = -\sum_{n=1}^{\infty} \frac{A^n}{n}, \tag{3.13}$$

then just as for real numbers, one sees this series converges absolutely for $\|A\| < 1$. Further, for all $B \in \mathscr{B}_1(1_V)$ one has

$$\exp(\log B) = B. \tag{3.14}$$

Formula (3.14) is known in the scalar case, and this implies that in fact (3.14) is an identity in absolutely convergent power series, whence it follows in the matrix case. The formulas (3.13) and (3.14) allow an alternate proof of Proposition 11 which avoids appeal to the Inverse Function Theorem and gives the explicit estimate that exp is 1-1 on $\mathscr{B}_{\log 2}(0)$. However, this explicit value of r is not needed, and we need in any case to appeal to the Inverse Function Theorem below in Theorem 17, so this more explicit proof of Proposition 11 gives us no particular benefit.

PROPOSITION 12. *Choose an $r < \log 2$, and let T be in $\exp \mathscr{B}_r(0)$, say $T = \exp A$. Then the transformation $S = \exp(A/2)$ is a square root of T; that is, $S^2 = T$. Moreover, S is the unique square root of T contained in $\exp \mathscr{B}_r(0)$.*

Proof. That $S^2 = T$ follows directly from Proposition 8. It is only necessary to prove the uniqueness of S. From Proposition 11, we see that our restriction on r implies $\exp \mathscr{B}_r(0) \subseteq \mathscr{B}_1(1_V)$. Hence it will suffice to show that if A, B are distinct linear maps of norm less than 1, then $(1_V + A)^2 \neq (1_V + B)^2$. Suppose the contrary. Then expanding the squares, cancelling the 1_V's and transposing, we find the equation

$$2(A - B) = B^2 - A^2 = B(B - A) + (B - A)A.$$

Taking norms yields

$$2\|A - B\| \leqslant \|B\|\,\|B - A\| + \|B - A\|\,\|A\| = (\|B\| + \|A\|)\|B - A\|.$$

This implies either $\|A - B\| = 0$, which is false since $A \neq B$, or $\|A\| + \|B\| \geqslant 2$, which is false since both $\|A\|$ and $\|B\|$ are less than 1. This contradiction establishes the uniqueness of S. ■

Proof of Theorem 10. Let $t \to M(t)$ be a continuous one-parameter group in $\mathrm{GL}(V)$. Since $M(0) = 1_V$, if we specify $r > 0$, we may by continuity and Proposition 11 find an $\varepsilon > 0$ such that $M(t) \in \exp(\mathscr{B}_r(0))$ for $|t| \leqslant \varepsilon$. We take $r < \log 2$. Write

$$M(\varepsilon) = \exp A_1$$

for appropriate $A_1 \in \mathscr{B}_r(0)$. If we set

$$A = \left(\frac{1}{\varepsilon}\right)A_1,$$

then $M(\varepsilon) = \exp(\varepsilon A)$. The transformations $M(\varepsilon/2)$ and $\exp((\varepsilon/2)A)$ are then both square roots of $M(t)$ lying in $\exp(\mathscr{B}_r(0))$. By Proposition 12 we conclude

$$M(\varepsilon/2) = \exp((\varepsilon/2)A).$$

An obvious induction using Proposition 12 shows that

$$M(2^{-n}\varepsilon) = \exp(2^{-n}\varepsilon A)$$

for all positive integers n. Taking mth powers, we conclude

$$M(m2^{-n}\varepsilon) = \exp(m2^{-n} \in A)$$

for all integers m and n. Since the numbers $m2^{-n}\varepsilon$ are dense in $\mathbb{R}$, Theorem 10 follows by continuity. ∎

4. Properties of the Exponential Map

The map exp is the basic link between the linear structure on End V and the multiplicative structure on $GL(V)$. We will describe some salient properties of this link.

Choose r with $0 < r \leqslant 1/2$ such that exp is one-to-one on $\mathscr{B}_r(0)$. Choose $r_1 < r$ so that if $A, B \in \mathscr{B}_{r_1}(0)$, then $\exp A \exp B$ is contained in $\exp \mathscr{B}_r(0)$. Then we can write

$$\exp A \exp B = \exp C \tag{4.1}$$

for some $C \in \mathscr{B}_r(0)$. The Inverse Function Theorem guarantees that C is a smooth (in fact analytic) function of A and B. There is a beautiful formula, the Campbell-Hausdorff formula [**J1**], [**Se**], which expresses C as a universal power series in A and B. To develop this completely would take too long. We will just give the first two terms in the expression for C. These suffice for most purposes.

For $A, B \in$ End V, write

$$[A, B] = AB - BA. \tag{4.2}$$

The quantity $[A, B]$ is called the *commutator* of A and B, and will be seen later to provide the Lie bracket operation in the Lie algebras we construct.

PROPOSITION 13. *Suppose A, B, C have norm at most $1/2$ and satisfy equation* (4.1). *Then we have*

$$C = A + B + \frac{1}{2}[A, B] + S, \tag{4.3}$$

where the remainder term S satisfies

$$\|S\| \leqslant 65(\|A\| + \|B\|)^3. \tag{4.4}$$

Proof. We have

$$\exp C = 1_V + C + R_1(C), \tag{4.5}$$

where the remainder $R_1(C)$ is

$$R_1(C) = \sum_{n=2}^{\infty} \frac{C^n}{n!}$$

and satisfies the obvious estimate

$$\|R_1(C)\| \leqslant \|C^2\| \left(\sum_{n=2}^{\infty} \frac{\|C\|^{n-2}}{n!} \right) \leqslant \|C\|^2$$

when $\|C\| \leqslant 1$, hence certainly when $\|C\| \leqslant 1/2$.

Similarly we have

$$\exp A \exp B = 1_V + A + B + R_1(A, B), \tag{4.6}$$

where by rearrangement of the double sum

$$R_1(A, B) = \sum_{n=2}^{\infty} \frac{1}{n!} \left(\sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \right).$$

Hence we have the estimate

$$\|R_1(A, B)\| \leqslant (\|A\| + \|B\|)^2 \left(\sum_{n=2}^{\infty} \frac{(\|A\| + \|B\|)^{n-2}}{n!} \right) \leqslant (\|A\| + \|B\|)^2$$

when $\|A\| + \|B\| \leqslant 1$.

Comparing equations (4.5) and (4.6), we see that equation (4.1) implies

$$C = A + B + R_1(A, B) - R_1(C). \tag{4.7}$$

Hence

$$\|C\| \leqslant \|A\| + \|B\| + (\|A\| + \|B\|)^2 + \|C\|^2 \leqslant 2(\|A\| + \|B\|) + \frac{1}{2}\|C\|$$

when A, B and C all have norm at most $\frac{1}{2}$.
Thus

$$\|C\| \leqslant 4(\|A\| + \|B\|). \tag{4.8}$$

Returning to equation (4.7), we further find

$$\begin{aligned} \|C - (A + B)\| \leqslant \|R_1(A, B)\| + \|R_1(C)\| &\leqslant (\|A\| + \|B\|)^2 + (4(\|A\| + \|B\|))^2 \\ &= 17(\|A\| + \|B\|)^2. \end{aligned} \tag{4.9}$$

We now refine these estimates to second order. In analogy with (4.5) we have

$$\exp C = 1_V + C + \frac{C^2}{2} + R_2(C), \tag{4.10}$$

where

$$R_2(C) = \sum_{n=3}^{\infty} \frac{C^n}{n!}$$

is easily estimated by

$$\|R_2(C)\| \leqslant \left(\frac{1}{3}\right)\|C\|^3 \tag{4.11}$$

when $\|C\| \leqslant 1$.

If we substitute expression (4.3) for C in equation (4.10), we obtain

$$\begin{aligned} \exp C &= 1_V + A + B + \frac{1}{2}[A, B] + S + \frac{1}{2}C^2 + R_2(C) \\ &= 1_V + A + B + \frac{1}{2}[A, B] + \frac{1}{2}(A + B)^2 + T \\ &= 1_V + A + B + \frac{1}{2}(A^2 + 2AB + B^2) + T, \end{aligned} \tag{4.12}$$

where

$$T = S + \frac{1}{2}\left(C^2 - (A + B)^2\right) + R_2(C).$$

On the other hand, we have

$$\exp A \exp B = 1_V + A + B + \frac{1}{2}(A^2 + 2AB + B^2) + R_2(A, B), \tag{4.13}$$

where

$$R_2(A, B) = \sum_{n=3}^{\infty} \frac{1}{n!}\left(\sum_{k=0}^{\infty} \binom{n}{k} A^k B^{n-k}\right)$$

satisfies $\|R_2(A, B)\| \leqslant \frac{1}{3}(\|A\| + \|B\|)^3$ when $\|A\| + \|B\| \leqslant 1$.

Comparison of (4.12) and (4.13) in the light of (4.1) yields

$$S = R_2(A, B) + \frac{1}{2}\left((A + B)^2 - C^2\right) - R_2(C).$$

Taking norms, we find

$$\begin{aligned}\|S\| &\leqslant \|R_2(A,B)\| + \frac{1}{2}(\|(A+B)(A+B-C)+(A+B-C)C\|) + \|R_2(C)\| \\ &\leqslant \frac{1}{3}(\|A\|+\|B\|)^3 + \frac{1}{2}(\|A\|+\|B\|+\|C\|)\|A+B-C\| + \frac{1}{3}\|C\|^3 \\ &\leqslant \frac{1}{3}(\|A\|+\|B\|)^3 + \frac{5}{2}(\|A\|+\|B\|)\cdot 17(\|A\|+\|B\|)^2 + \frac{1}{3}(4(\|A\|+\|B\|))^3 \\ &\leqslant 65(\|A\|+\|B\|)^3,\end{aligned}$$

as was to be shown.

We will derive two main consequences of Proposition 13. These relate group operations in $\mathrm{GL}(V)$ to the linear operations in $\mathrm{End}(V)$, and are crucial ingredients in the proof of the main theorem (Theorem 17 in §5) that relates Lie algebras to Lie groups. Proposition 14 relates group multiplication in $\mathrm{GL}(V)$ to addition in $\mathrm{End}(V)$, and Proposition 15 relates the group commutator operation to the bilinear commutator bracket defined in equation (4.2).

PROPOSITION 14 (Trotter Product Formula). *For A, $B \in \mathrm{End}\, V$, one has*

$$\exp(A+B) = \lim_{n\to\infty}(\exp(A/n)\exp(B/n))^n. \tag{4.14}$$

Proof. For n large enough, A/n and B/n will be close enough to the origin that formula (4.3) applies. We then have

$$\exp(A/n)\exp(B/n) = \exp C_n,$$

where by estimate (4.9)

$$\|C_n - (A+B)/n\| \leqslant 17((\|A\|+\|B\|)/n)^2.$$

Hence as $n \to \infty$, we see that $nC_n \to A+B$. Since $\exp nC_n = (\exp C_n)^n$, equation (4.14) follows. ■

Recall that the (linear) commutator $[A, B]$ is defined in equation (4.2). Recall also that if g, h are elements of a group, then the group commutator of g and h, written $(g:h)$, is the expression

$$(g:h) = ghg^{-1}h^{-1}.$$

PROPOSITION 15 (Commutator formula). *For A, $B \in \mathrm{End}\, V$, one has*

$$\begin{aligned}\exp[A,B] &= \lim_{n\to\infty}(\exp(A/n)\exp(B/n)\exp(-A/n)\exp(-B/n))^{n^2} \\ &= \lim_{n\to\infty}((\exp(A/n):\exp(B/n))^{n^2}.\end{aligned} \tag{4.15}$$

Proof. As in Proposition 14, for large n we have

$$\exp(A/n)\exp(B/n) = \exp C_n = \exp\left((A+B)/n + \frac{1}{2}\frac{[A,B]}{n^2} + S_n\right),$$

where

$$\|S_n\| \leqslant 65\frac{(\|A\|+\|B\|)^3}{n^3}.$$

Similarly

$$\exp(-A/n)\exp(-B/n) = \exp\left(-(A+B)/n + \left(\frac{1}{2}\right)\frac{[A,B]}{n^2} + S_n'\right) = \exp C_n'$$

with also

$$\|S_n'\| \leqslant 65\frac{(\|A\| + \|B\|)^3}{n^3}.$$

Hence

$$(\exp(A/n) : \exp(B/n)) = \operatorname{Exp} C_n \exp C_n' = \operatorname{Exp} E_n,$$

where

$$\begin{aligned} E_n &= C_n + C_n' + \frac{1}{2}[C_n, C_n'] + T_n \\ &= \frac{[A, B]}{n^2} + \frac{1}{2}[C_n, C_n'] + S_n + S_n' + T_n, \end{aligned}$$

where T_n is the term S in equation (4.3) if $A = C_n$ and $B = C_n'$.

It will suffice to show that there is a number γ, depending on A and B, such that

$$\left\| E_n - \frac{[A, B]}{n^2} \right\| \leqslant \frac{\gamma}{n^3}.$$

For then

$$(\exp E_n)^{n^2} = \exp([A, B] + U_n)$$

with $\|U_n\| \leqslant \gamma/n$, and equation (4.15) follows. In turn, it will suffice to show that the 2nd, 3rd, 4th and 5th terms in the expression for E_n are each less than a constant times n^{-3}. For S_n, S_n' and T_n, this follows from Proposition 13. Thus we need only worry about $[C_n, C_n']$. We compute

$$\begin{aligned} [C_n, C_n'] &= \left[\frac{1}{n}(A + B) + \frac{1}{2n^2}[A, B] + S_n, \frac{-1}{n}(A + B) + \frac{1}{2n^2}[A, B] + S_n'\right] \\ &= \frac{1}{n^3}[A + B, [A, B]] + \frac{1}{n}[A + B, S_n + S_n'] + \frac{1}{2n^2}[[A, B], S_n' - S_n] \\ &\quad + [S_n, S_n']. \end{aligned}$$

Using Proposition 13, we see that each of the four terms in this last sum is bounded by a constant times n^{-3}. (In fact, all terms except the first are bounded by a constant times n^{-4}).

There is one further concept involving the exponential map that is basic to Lie theory. It involves conjugation, which is generally referred to as the "adjoint action." For $g \in \mathrm{GL}(V)$ and $T \in \operatorname{End} V$, we can form the conjugate

$$\operatorname{Ad} g(A) = gAg^{-1}. \tag{4.16}$$

The following proposition is easily verified and left as an exercise.

PROPOSITION 16. (i) $\operatorname{Ad} g(aA + bB) = a\operatorname{Ad} g(A) + b\operatorname{Ad} g(B)$ *for* $A, B \in \operatorname{End} V$; $a, b \in \mathbb{R}$; *and* $g \in \mathrm{GL}(V)$.

(ii) $\operatorname{Ad} g(AB) = \operatorname{Ad} g(A)\operatorname{Ad} g(B)$.

(iii) $\operatorname{Ad} g_1 g_2(A) = \operatorname{Ad} g_1(\operatorname{Ad} g_2(A))$.

Formulas (i) and (ii) say $\operatorname{Ad} g$ is an algebra automorphism of $\operatorname{End} V$, and Formula (iii) says the map $\operatorname{Ad}: g \to \operatorname{Ad} g$ is a group homomorphism from $\mathrm{GL}(V)$ to the automorphism group of $\operatorname{End}(V)$. The map Ad is called the *adjoint action* of $\mathrm{GL}(V)$ on $\operatorname{End}(V)$.

Formula (iii) implies in particular that if $\exp tA$ is a one-parameter subgroup of $\mathrm{GL}(V)$, then $\operatorname{Ad} \exp tA$ is a one-parameter group of linear transformations on $\operatorname{End} V$. Hence $\operatorname{Ad} \exp tA$ has infinitesimal generator $\mathscr{A} \in \operatorname{End}(\operatorname{End} V)$. We can compute $\mathscr{A}$ by the formula

$$\begin{aligned}\mathscr{A}(B) &= \lim_{t\to 0}\frac{(\exp tA)B(\exp(-tA)) - B}{t}\\ &= \frac{d}{dt}(\exp tA)B(\exp(-tA))\big|_{t=0}\\ &= (A(\exp tA)B(\exp - tA) + (\exp tA)B(-A)(\exp - tA))\big|_{t=0}\\ &= AB - BA = [A,B].\end{aligned}$$

Here we have used the fact that

$$\frac{d}{dt}(\exp(tA)) = A\exp(tA).$$

This formula may be verified by direct calculation from the definition of $\exp(tA)$. Hence if we define

$$\operatorname{ad}A : \operatorname{End} V \to \operatorname{End} V$$

by

$$\operatorname{ad}A(B) = [A,B],$$

we have the following formula.

PROPOSITION 17. *For* $A \in \operatorname{End} V$

$$\operatorname{Ad}(\exp A) = \exp(\operatorname{ad}A). \tag{4.17}$$

5. The Lie Algebra of a Matrix Group

By a *matrix group* we mean a closed subgroup of $\mathrm{GL}(V)$ for some vector space V. This section shows a matrix group is a Lie group. What that means is expressed in Theorem 17. Most, though not all, Lie groups can be realized as matrix groups. This article discusses only matrix groups.

EXAMPLES. (i) $\mathrm{GL}(V)$ itself.

(ii) $\mathrm{SL}_n(\mathbb{R})$, the special linear group, of $n \times n$ matrices of determinant 1.

(iii) $O_{p,q}$, the "pseudo-orthogonal groups," consisting of all matrices in $\mathrm{GL}_{p+q}(\mathbb{R})$ that preserve the indefinite inner product

$$(x,x')_{p,q} = \sum_{i=1}^{P} x_i x_i' - \sum_{i=p+1}^{p+q} x_i x_i',\ x, x' \in \mathbb{R}^{p+q}.$$

(iv) $\mathrm{SP}_{2n}(\mathbb{R})$, the real symplectic group, consisting of all matrices in $\mathrm{SL}_{2n}(\mathbb{R})$ that preserve the skew-symmetric bilinear form

$$\langle x, x'\rangle = \sum_{i=1}^{n} x_i x_{i+n}' - x_i' x_{i+n} x,\ x' \in \mathbb{R}^{2n}.$$

(v) The group $P(U)$ of transformations that preserve a subspace U of V. For instance, if $V = \mathbb{R}^n$, and $U_m = \mathbb{R}^m = \{(x_1, x_2, \dots, x_m, 0, 0, \dots, 0)\}$, where $m \leqslant n$, then

$$P(U_m) = \left\{\begin{bmatrix} A & X\\ 0 & B\end{bmatrix} : A \in \mathrm{GL}_m(\mathbb{R}),\ B \in \mathrm{GL}_{n-m}(\mathbb{R}),\ X \in M_{m,n-m}(\mathbb{R})\right\}.$$

Here $M_{m,n-m}(\mathbb{R})$ is the space of $m \times (n-m)$ real matrices.

(vi) Any intersection of matrix groups is a matrix group. For instance, the intersection $\cap_{m=1}^{n} P(U_m)$ of the groups $P(U_m)$ of example (v) is the group of invertible upper triangular matrices.

(vii) The group preserving some closed subgroup, not necessarily a subspace, of V. For example, let $\mathbb{Z}^n \subseteq \mathbb{R}^n$ be the discrete subgroup of vectors with integral entries. Set

$$\mathrm{GL}_n(\mathbb{Z}) = \{A \in \mathrm{GL}_n(\mathbb{R}) : A(\mathbb{Z}^n) = \mathbb{Z}^n\}.$$

Then $\mathrm{GL}_n(\mathbb{Z})$ can also be shown to consist of matrices with integer entries and determinant ± 1.

(viii) The group commuting with some family $\{T_i\}$ of operators on V is a matrix group. For example, we can identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ by letting x_{2j-1} and x_{2j} be the real and imaginary parts of the coordinate z_j of $z = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$. If we do so, the operation of multiplication by a complex scalar becomes some (real) linear operator on $\mathbb{R}^{2n}$. Further, the group $\mathrm{GL}_n(\mathbb{C})$ becomes identified with the subgroup of $\mathrm{GL}_{2n}(\mathbb{R})$ formed by elements which commute with the multiplications by complex scalars.

(ix) If G_i is a matrix group in $\mathrm{GL}(V_i)$, $i = 1, 2$, then $G_1 \times G_2$ is a matrix group in $\mathrm{GL}(V_1 \oplus V_2)$ in the obvious way.

(x) If G is a matrix group, then G^0, the connected component of the identity in G, is a matrix group.

(xi) The normalizer in $\mathrm{GL}(V)$ of a matrix group is a matrix group.

The main result of this section is the essential phenomenon behind Lie theory: a matrix group has naturally attached to it a Lie algebra. Before showing this we recall what a Lie algebra is.

DEFINITION. A real *Lie algebra* $\mathfrak{g}$ is a real vector space equipped with a product

$$[\,,]: \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{5.1}$$

satisfying the identities

(5.2)

(i) (Bilinearity). For $a, b \in \mathbb{R}$ and $x, y, z \in \mathfrak{g}$,

$$[ax + by, z] = a[x, z] + b[y, z]$$

$$[z, ax + by] = a[z, x] + b[z, y].$$

(ii) (Skew symmetry). For $x, y \in \mathfrak{g}$,

$$[x, y] = -[y, x].$$

(iii) (Jacobi Identity). For $x, y, z \in \mathfrak{g}$,

$$[x,[y, z]] + [z,[x, y]] + [y,[z, x]] = 0.$$

The first main example of a Lie algebra is End V equipped with the bracket operation $[\,,]$ of commutator, as given in equation (4.2). It is left as an exercise to verify that this satisfies the correct identities. Any subspace of End V which is closed under $[\,,]$ will become a Lie algebra in its own right. Since our main theorem will provide us with such a subspace for each matrix group, we will postpone a more explicit discussion of examples.

Consider a matrix group $G \subseteq \mathrm{GL}(V)$. Let $\exp^{-1}(G) \subseteq \mathrm{End}\, V$ be the inverse image of G under exp. Since $\exp(nA) = (\exp A)^n$, it is clear that $\exp^{-1}(G)$ is closed under scalar multiplication by integers. Set

$$\mathfrak{g} = \{A \in \mathrm{End}\, V : \exp tA \in G \text{ for all } t \in \mathbb{R}\} = \bigcap_{t \in \mathbb{R}^{\times}} t \exp^{-1}(G).$$

Observe that $\mathfrak{g}$ is the collection of infinitesimal generators of one-parameter subgroups of G. We call $\mathfrak{g}$ the *Lie algebra* of G.

THEOREM 17. (a) *The Lie algebra* $\mathfrak{g}$ *of a matrix group* G *is a Lie algebra.*

(b) *The map* $\exp : \mathfrak{g} \to G$ *maps a neighborhood of* 0 *in* $\mathfrak{g}$ *bijectively onto a neighborhood of* 1_V *in* G.

REMARKS. (i) Part (b) of Theorem 17 implies G is locally homeomorphic to Euclidean space. In

fact it is not hard to refine part (b) and show that G has the structure of a smooth manifold, such that the group multiplication is smooth, but we will not do that here.

(ii) Theorem 17 provides a geometric picture of the relation between $\mathfrak{g}$ and G. If a one-parameter group $\exp(tA)$ is regarded as a curve inside the vector space End V, then this curve passes through the identity 1_V at time $t = 0$. By differentiating the formula for exp tA, we see the tangent vector at the point 1_V to this curve is just A. Thus, as we have defined it, $\mathfrak{g}$ consists simply of all tangent vectors to the curves defined by one-parameter groups in G. But Theorem 17 asserts that these tangent vectors actually fill out some linear subspace (namely $\mathfrak{g}$) of End V, and further, if we make the smooth change of coordinates $A \to \exp A$, then this linear subspace $\mathfrak{g}$ is bent in such a way that it lies entirely in G, and fills up G around 1_V. In other words, G is shown to be a smooth multidimensional surface inside End V, and $\mathfrak{g}$ is simply its tangent space at the point 1_V.

The main burden of the proof of Theorem 17 is carried by the following technical result.

LEMMA 18. *Suppose* $\{A_n\}$ *is a sequence in* $\exp^{-1}(G)$, *and* $\|A_n\| \to 0$. *Let* s_n *be a sequence of real numbers. Then any cluster point of* s_nA_n *is in* $\mathfrak{g}$.

Proof. Let B be the cluster point. By passing to a subsequence if necessary we may assume that s_nA_n converges to B. Fix a number $t \in \mathbb{R}$. Let m_n be an integer such that $|m_n - ts_n| \leqslant 1$. Then m_nA_n converges to tB; for we have

$$\begin{aligned}\|m_nA_n - tB\| &= \|(m_n - ts_n)A_n + t(s_nA_n - B)\|\\ &\leqslant |m_n - ts_n|\,\|A_n\| + |t|\,\|s_nA_n - B\|\\ &\leqslant \|A_n\| + |t|\,\|s_nA_n - B\|\end{aligned}$$

which converges to zero as $n \to \infty$, by our assumptions on A_n and B. Since $m_nA_n \in \exp^{-1}(G)$, and $\exp^{-1}(G)$ is closed, we see that $tB \in \exp^{-1}(G)$. Since t was arbitrary in $\mathbb{R}$, we see that $B \in \mathfrak{g}$. ■

Proof of Theorem 17. We first show $\mathfrak{g}$ is a subspace of End V. Since $\mathfrak{g}$ is by definition closed under scalar multiplication, we need only show it is closed under addition. Take $A, B \in \mathfrak{g}$. Then as in Proposition 14 we know that for large enough n

$$\exp(A/n)\exp(B/n) = \exp C_n,$$

where $\|C_n\| \to 0$, and $nC_n \to A + B$. Hence Lemma 18 implies $A + B \in \mathfrak{g}$.

Next we show that if $A, B \in \mathfrak{g}$, then also $[A, B] \in \mathfrak{g}$. As in Proposition 15 we know that for large n we have

$$(\exp(A/n) : \exp(B/n)) = \exp E_n$$

with $E_n \to 0$ and $n^2E_n \to [A, B]$. Another application of Lemma 18 says $[A, B] \in \mathfrak{g}$. This concludes part (a) of Theorem 17.

We know $\mathfrak{g}$ is a linear subspace of End(V). Let $Y \subseteq$ End(V) be a complementary subspace of $\mathfrak{g}$, so that End $V = \mathfrak{g} \oplus Y$. Let p_1 and p_2 be the projections of End V on $\mathfrak{g}$ and Y, respectively, with respective kernels Y and $\mathfrak{g}$. Define a map E : End $V \to$ GL(V) by

$$E(A) = \exp(p_1(A))\exp(p_2(A)).$$

By use of Proposition 13, we can compute that

$$\frac{d}{dt}\left(\exp(p_1(tA))\exp(p_2(tA))\right)\Big|_{t=0} = p_1(A) + p_2(A) = A.$$

This says that the differential of E at 0 is the identity map on End V, so that E takes small neighborhoods of 0 to neighborhoods of 1_V bijectively, by the Inverse Function Theorem. Choose a small ball $\mathcal{B}_r(0) \subseteq$ End V, and suppose $\exp(\mathcal{B}_r(0) \cap \mathfrak{g})$ does not cover a neighborhood of 1_V in G. Then we can find a sequence $B_n \in \exp^{-1}(G)$ such that $B_n \to 0$, but $B_n \notin \mathfrak{g}$. When B_n is close enough to 0, we may write

$$\exp B_n = E(A_n)$$

for some A_n. We will have $A_n \to 0$ as $B_n \to 0$. Then

$$\exp(p_2(A_n)) = \exp(p_1(A_n))^{-1} \exp B_n$$

is also in G, and is nonzero by our assumption on B_n. Since $A_n \to 0$, $p_2(A_n) \to 0$ also. The sequence $\|p_2(A_n)\|^{-1} p_2(A_n)$ will have cluster points, and these must be in $\mathfrak{g}$ by Lemma 18. On the other hand, $p_2(A_n) \in Y$, so all cluster points must be in Y. This contradicts the fact that Y was chosen complementary to $\mathfrak{g}$, so statement (b) of Theorem 17 follows. ∎

EXAMPLES. We will describe below the Lie algebras of some of the groups listed at the beginning of this section. The verification that the indicated Lie algebras are indeed the Lie algebras of the stated groups is left as an exercise.

(i) The Lie algebra of $\mathrm{GL}(V)$ is of course $\mathrm{End}(V)$.

(ii) The Lie algebra of $\mathrm{SL}_n(\mathbf{R})$ is the space of $\mathrm{sl}_n(\mathbf{R})$ of $n \times n$ matrices of trace zero.

(iii) Let β be a bilinear form on V. The *isometry group of* β is the group of invertible operators A such that

$$\beta(Au, Av) = \beta(u, v) \qquad \text{for all} \qquad u, v \in V.$$

The Lie algebra of this group is the space of operators B such that

$$\beta(Bu, v) + \beta(u, Bv) = 0.$$

In particular the Lie algebra $\mathfrak{o}_n(V)$ of the orthogonal group $O_n(\mathbf{R})$ of isometries of the standard inner product on $\mathbf{R}^n$ is the space of skew-symmetric matrices.

(iv) The Lie algebra of the subgroup of $\mathrm{GL}(V)$ of maps commuting with given operators $\{T_i\}$ is the subalgebra of End V commuting with the T_i.

(v) The Lie algebra of the group $P(\{V_i\})$ of invertible transformations which preserve each of the subspaces V_i of V is the subalgebra of all transformations which preserve the V_i. In particular, the Lie algebra of the group of invertible upper triangular matrices is the vector space of all upper triangular matrices.

(vi) The Lie algebra of $G_1 \cap G_2$, for matrix groups G_i, is $\mathfrak{g}_1 \cap \mathfrak{g}_2$.

(vii) A matrix group G and its identity component G^0 have the same Lie algebra.

After its existence, the second most important feature of $\mathfrak{g}$ is that it is natural (in the sense of category theory). This is the content of our next theorem.

Let $\mathfrak{g}$, $\mathfrak{h}$, be real Lie algebras. A *homomorphism* from $\mathfrak{g}$ to $\mathfrak{h}$ is a linear map

$$L: \mathfrak{g} \to \mathfrak{h}$$

satisfying

$$L([x, y]) = [Lx, Ly] \qquad x, y \in \mathfrak{g}. \tag{5.3}$$

Let V, U be real vector spaces.

THEOREM 19. *Let $G \subseteq \mathrm{GL}(V)$ be a matrix group with Lie algebra $\mathfrak{g}$. Let $\phi: G \to \mathrm{GL}(U)$ be a continuous homomorphism. Then there is a homomorphism of Lie algebras*

$$d\phi: \mathfrak{g} \to \mathrm{End}\, U \tag{5.4}$$

such that

$$\exp(d\phi(A)) = \phi(\exp A). \tag{5.5}$$

Proof. If $A \in \mathfrak{g}$, then $\exp tA$ is a one-parameter subgroup of G, so $\phi(\exp(tA))$ is a one-parameter subgroup of $\phi(G) \subseteq \mathrm{GL}(U)$. Hence by Theorem 10 we may write $\phi(\exp(tA)) = \exp(tB)$ for some $B \in \mathrm{End}\, U$. If we define

$$d\phi(A) = B,$$

then equation (5.5) will obviously be satisfied. To prove this theorem, it suffices to show that $d\phi$ is a homomorphism of Lie algebras. But this follows directly from Propositions 14 and 15 which show that the Lie algebra operations in $\mathfrak{g}$ are determined by operations in G.

EXAMPLE. The formula (4.17) shows that $\mathfrak{g}$ is an invariant subspace of $\operatorname{End} V$ under the operators $\operatorname{Ad} g$, $g \in G$. The restriction of $\operatorname{Ad} g$ to $\mathfrak{g}$ is again denoted by $\operatorname{Ad} g$, and the resulting action of G on $\mathfrak{g}$ is still called the adjoint action. In terms of Theorem 19, the formula (4.17) has the interpretation

$$d(\mathrm{Ad}) = \mathrm{ad}. \tag{5.6}$$

An immediate consequence of Theorem 19 is:

COROLLARY 20. *If $G_1 \subseteq \mathrm{GL}(V)$ and $G_2 \subseteq \mathrm{GL}(U)$ are isomorphic matrix groups, then their Lie algebras $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are isomorphic as Lie algebras.*

Proof. Let $\phi : G_1 \to G_2$ be a continuous isomorphism with continuous inverse ϕ^{-1}. Then in particular ϕ is a continuous homomorphism from G_1 to $\mathrm{GL}(U)$, and ϕ^{-1} is a continuous homomorphism from G_2 to $\mathrm{GL}(V)$. Theorem 19 therefore provides us with associated Lie algebra homomorphisms $d\phi$ and $d(\phi^{-1})$. It follows from the definition of the Lie algebra of a matrix group and formula (5.5) that in fact $d\phi(\mathfrak{g}_1) \subseteq \mathfrak{g}_2$, and similarly $d(\phi^{-1})(\mathfrak{g}_2) \subseteq \mathfrak{g}_1$. It further follows from formula (5.5) that since $\phi^{-1} \circ \phi$ is the identity on G_1, then also $d(\phi^{-1}) \circ d\phi$ is the identity on $\mathfrak{g}_1$. In other words $d(\phi^{-1}) = (d\phi)^{-1}$, so $d\phi$ is in fact a Lie algebra isomorphism from $\mathfrak{g}_1$ to $\mathfrak{g}_2$. ■

The converse of Corollary 20, that groups with isomorphic Lie algebras are isomorphic, is false. For example the rotation group

$$\mathrm{SO}_2 = \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} : \theta \in \mathbb{R} \right\}$$

and the diagonal group

$$D_1 = \left\{ \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} : a \in \mathbb{R}, \quad a > 0 \right\}$$

both have Lie algebra isomorphic to $\mathbb{R}$, but SO_2 is homeomorphic to a circle, while D_1 is homeomorphic to $\mathbb{R}$, so they are certainly not isomorphic.

However, the converse of Corollary 20 is in a sense almost true, so that the bracket operation on $\mathfrak{g}$ almost determines G as a group. After the existence of the Lie algebra, this fact is the most remarkable in Lie theory. Its precise formulation is known as Lie's Third Theorem. It is in proving a suitable version of Lie's Third Theorem that Lie theory begins to get involved, so we will leave the story here. Precise treatments of these issues can be found in **[A]**, **[Ch]**, **[He]**, **[Se]**.

6. Loose Ends and Further Developments

In §§3, 4, and 5 we have shown that to each matrix group G, there is associated in a close and natural way, a Lie algebra $\mathfrak{g}$, the two being connected via one parameter groups and the exponential map. These facts constitute an important part of the foundations of Lie theory. We will describe briefly what we have omitted from the standard account.

First, we have not treated Lie groups as abstract things-in-themselves, but have only dealt with them as subgroups of a standard group, $\mathrm{GL}(V)$. We could not have discussed abstract Lie groups without assuming the standard language of differentiable manifolds. Our approach allowed us to bring to the fore the remarkable Theorem 17, which asserts that merely the requirements of being closed and being a group inside $\mathrm{GL}(V)$ (or any Lie group) suffices to make the group a smooth manifold. This indicates what a strong regularity condition the group property is. Research over the past decades have continued to underscore this theme **[BT]**, **[Ma]**, **[Mo]**.

Second, we have not demonstrated how complete and mutual is the relationship between Lie groups G and their Lie algebras $\mathfrak{g}$. It is in this direction that the principal technical complications of the theory lie. For example, although we have shown how to attach a Lie algebra to every matrix group, we have not tried to attach a group to every Lie subalgebra of End V. Indeed, this is not possible if one sticks to matrix groups; the one parameter groups obtained by exponentiating elements in a given Lie algebra $\mathfrak{g}$ will generate a group which in a suitable sense has $\mathfrak{g}$ as its Lie algebra but this group will not always be closed in $\mathrm{GL}(V)$. The simplest example is probably the one-parameter group $\exp tA_x$ in $\mathrm{GL}_4(\mathbf{R})$, where

$$A_x = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x \\ 0 & 0 & -x & 0 \end{bmatrix}$$

and x is any irrational number. Also the question of the relation of two matrix groups which have isomorphic Lie algebras, essentially the question of the converse of Corollary 20, involves the notion of covering space and fundamental group **[Ms]** and is beyond the scope of this discussion. Interestingly enough, both these questions are most vexed for the most simple-minded case: abelian Lie groups and their Lie algebras. We close these brief remarks by pointing out that, when G is fairly nonabelian, especially if the center of G is discrete, the existence of the adjoint action and formula (5.6) in particular go a long way toward showing that G is nearly determined by $\mathfrak{g}$. After the foundations comes the rather extensive development of the structure theory of Lie algebras, with direct consequences for the groups. Several fine accounts of the theory of Lie algebras are available, for example **[J]**, **[Hu]**. Beyond the theory of Lie groups and algebras in themselves lies the vast domain of their applications. We have mentioned a few of these in the introduction and in §7. Some representative references for applications are **[BC]**, **[HP]**, **[Hr]**, **[Ko]**, **[Lo]**.

Our treatment in §§3, 4, 5 has been concrete in that we worked only inside End V, but it was also abstract in that it was coordinate free. We record here some common terminology used when bases are introduced. Let $\mathfrak{g} \subseteq \mathrm{End}\, V$ be a Lie subalgebra. Let $\{y_i\}, 1 \leqslant i \leqslant \dim \mathfrak{g}$ be a basis for $\mathfrak{g}$. Then the fact that $\mathfrak{g}$ is a Lie algebra amounts to the statement that the commutators $[y_i, y_j]$ are again linear combinations of the y_k's. Thus we have equations

$$[y_i, y_j] = \sum c_{ij}^k y_k, \tag{6.1}$$

where the c_{ij}^k are real numbers. The equations (6.1) are called the *commutation relations* of the y_i's and the c_{ij}^k are called the *structure constants* of $\mathfrak{g}$ with respect to the y_i.

For example, set

$$e^+ = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad e^- = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad h = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The matrices e^+, e^-, and h form a basis for sl_2, the 2×2 traceless matrices, one of the most fundamental Lie algebras. It is easy to compute that their commutation relations are

$$[h, e^+] = 2e^+ [h, e^-] = -2e^- [e^+, e^-] = h.$$

7. Relations with the Standard Curriculum

In this section we give some examples of how Lie theory makes contact with current staples of undergraduate mathematics. We must of course be very restrictive and brief.

1. Many of the standard theorems of linear algebra are of course also part of the fabric of Lie theory, and gain coherence when considered in that light. For example, several of the standard canonical forms, e.g., Jordan form, the diagonalization of the (skew) Hermitian matrices, amount to classification of the conjugacy classes (orbits under the adjoint action) in a Lie algebra. Jordan form describes conjugacy classes in $\mathrm{End}(\mathbf{C}^n) \sim \mathfrak{gl}_n(\mathbf{C})$, and diagonalization of Hermitian matrices

describes conjugacy classes in U_n, the $n \times n$ unitary group. We cannot explain this interpretation of these results in detail, but encourage the reader to explore it by further reading.

Also, $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}_n(\mathbb{C})$ are examples of an extremely important class of Lie groups called semisimple groups, and several well-known results in linear algebra are special cases for $\mathrm{SL}_n(\mathbb{R})$ or $\mathrm{SL}_n(\mathbb{C})$ of structure theorems for semisimple groups. (Since GL_n and SL_n are so similar, we state the results for GL_n.) The polar decomposition or singular value decomposition **[St]** says that any $A \in \mathrm{GL}_n(\mathbb{R})$ may be written in the form

$$A = OS = O_1 D O_2,$$

where O, O_1, and O_2 are orthogonal matrices, S is symmetric, and D is diagonal with positive entries. This is the specialization to $\mathrm{GL}_n(\mathbb{R})$ of what is known as the *Cartan decomposition* **[He]** in the context of semisimple Lie groups. Also, the Gram-Schmidt orthonormalization procedure **[St]** says, in group-theoretical terms, that any $A \in \mathrm{GL}_n(\mathbb{R})$ may be written in the form

$$A = OB = ODU,$$

where O is orthogonal, B is upper triangular, D is diagonal with positive entries, and U is upper triangular with diagonal entries all equal to 1. For general semisimple groups, this is known as the *Iwasawa decomposition* **[He]**.

Various basic features in the elimination theory, including the "LU factorization" **[St]** of a generic matrix into the product of an upper triangular and a lower triangular matrix, and the "reduced row-echelon form" **[DN]** are aspects of a different kind of decomposition of semisimple groups, known as the *Bruhat decomposition* **[Bo]**.

2. The cross product on $\mathbb{R}^3$ defines a Lie algebra structure on $\mathbb{R}^3$. This is in fact isomorphic to $\mathfrak{o}_3$, the Lie algebra of O_3, the 3×3 skew symmetric matrices. The isomorphism is accomplished by

$$(x, y, z) \to \begin{bmatrix} 0 & -x & -y \\ x & 0 & -z \\ y & z & 0 \end{bmatrix}.$$

The generalization of this correspondence to higher dimensions leads to the theory of spinors and Clifford algebras **[J2]**.

3. The fact that second mixed partial derivatives are equal is a reflection of the fact that $\mathbb{R}^n$ is an abelian Lie group.

4. The theory of Fourier series and Fourier transform is best understood group-theoretically. See **[Gr]** for a discussion.

5. It is fairly routine in quantum mechanics courses, in conjunction with the Schrödinger equation for the hydrogen atom and angular momentum, to introduce, "raising and lowering operators" **[Me]**. The operators belong to the complexification of $\mathfrak{o}_3$, which is isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. The commutation relations of the Lie algebra figure importantly in the computations. The harmonic oscillator is also susceptible to a Lie-theoretic treatment. The Canonical Commutation Relations themselves are the laws for a bracket relation on a Lie algebra, known as the Heisenberg Lie algebra **[Ca]**, **[Ho]**. The relations of this algebra with quantum mechanics, and physics generally, is deep and extensive.

6. Perhaps the part of standard undergraduate mathematics that is pedagogically most compatible with Lie theory is differential equations. We have already discussed in §2 how the notion of one-parameter group is a geometrization of the solution of a system of differential equations. And in §3 we noted that one-parameter groups of linear transformations were associated with the very important class of linear, constant coefficient systems. Indeed, the exponential map and linear algebra techniques are often explicitly used in treating these systems **[Br]**.

Many of the important classical differential equations are related with Lie theory. Indeed much of the theory of special functions may be considered a branch of Lie theory [**Mi**], [**V**]. Below I state, always by way of example, some exercises which I have given to students in differential equations courses and which were favorably received.

A.(i) Let P, Q and the identity operator I span a Lie algebra, with commutation relations $[P, Q] = I$, and of course $[P, I] = [Q, I] = 0$. (These are the Canonical Commutation Relations.) Define $L = (P - I)QP$, and $A_n = (P - I)^n Q^n$ (so $L = A_1 P$). Show that

$$\begin{aligned}
&(a) \quad [Q,(P-I)^n] = -n(P-I)^{n-1},\\
&(b) \quad A_{n+1} = (A_1 + n)A_n,\\
&(c) \quad [L, A_1] = L - A_1, \text{ and}\\
&(d) \quad L(A_1 + n) = (A_1 + n)L + (L + n) - (A_1 + n).
\end{aligned}$$

(ii) Suppose v_n is an eigenvector for L, with eigenvalue $-n$, so that $(L + n)v_n = 0$. Show from (d) above that $(A_1 + n)v_n$ is an eigenvector for L, with eigenvalue $-(n + 1)$. Conclude from (b) that if v_0 is an eigenvector of L with eigenvalue 0, then $A_n v_0 = v_n$ is an eigenvector with eigenvalue $-n$.

(iii) Show that if $P = d/dx$ and $Q =$ multiplication by x, then P and Q satisfy the relations above. Show also that

$$e^x \frac{d}{dx} e^{-x} = P - I.$$

Conclude that a solution to the Laguerre equation $zy'' + (1 - z)y' + ny = 0$ is

$$e^x\left(\frac{d}{dx}\right)^n (e^{-x}x^n) = (P - I)^n Q^n(1);$$

here 1 is the constant function on $\mathbb{R}$.

B.(i) Take P, Q and I as in A(i). Suppose $Pv_0 = 0$, and set $v_n = Q^n(v_0)$. Show inductively that $Pv_n = nv_{n-1}$. Conclude that v_n is an eigenvector of eigenvalue n for QP.

(ii) Put $P = d/dx, Q = (d/dx) + x$. Verify that these satisfy the correct commutation relations, and show that

$$Q = e^{-x^2/2} \frac{d}{dx} e^{x^2/2}.$$

Show that solutions of Hermite's equation $y'' + xy' - ny = 0$ are given by

$$y = e^{-x^2/2}\left(\frac{d}{dx}\right)^n e^{x^2/2} = Q^n(1).$$

In addition to Rodrigues-type formulas such as the above, one can deduce in a purely formal manner recursion relations and other properties of the Hermite, Laguerre, Legendre, Bessel, and many other classical families of functions.

I would like to thank Kenneth Gross for painstaking efforts to improve the readability of this paper.

References

[**A**] J. F. Adams, Lectures on Lie Groups, Benjamin, New York, 1969.

[**Be**] L. Bers, Riemann Surfaces, New York University Lecture Notes, 1957–58.

[**Bo**] A. Borel, Linear Algebraic Groups, Benjamin, New York, 1969.

[**Bo2**] A. Borel, On the development of Lie Group Theory, Math. Intelligencer, 2 (1980) 67–72.

[**BC**] A. Borel and W. Casselman, Editors, Automorphic Forms, Representations, and L-functions, Proc. Symp. Pure Math. XXXIII, American Mathematical Society, Providence, R.I., 1979.

[**BT**] A. Borel and J. Tits, Homomorphismes "abstraits" de groupes algébriques simples, Ann. of Math., (2) 97 (1973) 499–571.

[Br] M. Braun, Differential Equations and Their Applications, 2nd ed., Springer Verlag, New York, 1978.

[Ca] P. Cartier, Quantum mechanical commutation relations and theta functions, Proc. Symp. Pure Math. IX, American Mathematical Society, Providence, R.I., 1966, 361–383.

[Ch] C. Chevalley, Theory of Lie Groups I, Princeton University Press, Princeton, N.J., 1946.

[DN] J. Daniel, B. Noble, Applied Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1977.

[Dy] F. J. Dyson, Mathematics in the Physical Sciences, Scientific American, 211(Sept. 1969) 128–147.

[E] A. Einstein, Zur Elektrodynamik bewegter Korper, Ann. Physik, Bd. 17 (1905) 891.

[F] R. Fenn, What is the geometry of a surface?, this MONTHLY, 90 (1983) 87–98.

[G] A. Gleason, Groups without small subgroups, Ann. of Math., 56 (1952) 193–212.

[Go] R. Godement, Introduction à la théorie des groupes de Lie (in 2 volumes), Pub. Math. de l'Univ. Paris VII, vol. 11, vol. 12, Paris, 1982.

[Gr] K. Gross, On the evolution of noncommutative harmonic analysis, this MONTHLY, 85 (1978) 525–528.

[Hg] W. Heisenberg, Uber quantentheoretische Umdeutung kinematischer und mechanischen Beziehungen, Phys., 33 (1925) 879.

[He] S. Helgason, Differential Geometry, Lie Groups, and Symmetric Spaces, Academic Press, New York, 1978.

[HP] E. Hille and R. Phillips, Functional Analysis and Semigroups, A.M.S. Colloquium Publications, vol. 31, American Mathematical Society, Providence, R.I., 1957.

[HS] M. Hirsch and S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, New York, 1974.

[Ho] R. Howe, On the role of the Heisenberg group in harmonic analysis, B.A.M.S. (New Series), 3 (1980) 821–843.

[Hu] J. Humphreys, Introduction to Lie algebras and representation theory, Springer Verlag, New York, 1972.

[Hn] T. Husain, Introduction to Topological Groups, Saunders, PA, 1966.

[Hr] D. Husemoller, Fibre Bundles, McGraw-Hill, New York, 1966.

[J1] N. Jacobson, Lie Algebras, Wiley-Interscience, New York, 1962.

[J2] N. Jacobson, Basic Algebra II, Freeman, San Francisco, 1980.

[Ka] I. Kaplansky, Lie Algebras and Locally Compact Groups, Chicago University Press, Chicago, 1971.

[Ke] J. Kelley, General Topology, Van Nostrand, Princeton, N.J., 1955.

[Kl] F. Klein, Vergleichende Betractungen uber neuere geometrische Forschungen, in Mathematische Abhandlungen, 1st ed., Springer-Verlag, Berlin, 1921.

[Ko] S. Kobayashi, Transformation Groups in Differential Geometry, Ergeb. Math., b. 70, Springer-Verlag, New York, Heidelberg, 1970.

[L] S. Lang, Analysis I, Addison-Wesley, Reading, MA, 1968.

[Lo] E. Loebl, Editor, Group Theory and Its Applications I, II, Academic Press, New York, 1968, 1971.

[Ma] G. Margulis, On the arithmeticity of nonuniform lattices, Proc. of Int. Cong. Math., Vancouver, 1970, pp. 21–35. (Russian).

[Ms] W. Massey, Algebraic Topology: An Introduction, Graduate Texts in Math., 56, Springer-Verlag, New York, Heidelberg, 1967.

[Me] A. Messiah, Quantum Mechanics II, North-Holland, Amsterdam, 1963.

[Mi] W. Miller, Lie theory and Special Functions, Academic Press, New York, 1968.

[Mk] H. Minkowski, Die Grundgleichungen fur die elektromagnetischen Vorgange in Bewegten Korpern, Nachr. d. K. Ges. d. Wissensch. zu Gottingen, 1908, 53.

[MZ] D. Montgomery and L. Zippin, Topological Transformation Groups, Wiley Interscience, New York, 1955.

[Mo] G. D. Mostow, Strong rigidity of locally symmetric spaces, Ann. of Math. Studies, no. 78, Princeton University Press, Princeton, N.J., 1973.

[N] M. Naimark, Normed Algebras, P. Noordhoff, N. V., Groningen, Netherlands, 1959.

[Nn] J. von Neumann, Uber die analytischen Eigenschaften Gruppen linearer Transformationen und ihrer Darstellungen, Math. Z., 30 (1929) 3–42.

[P] L. S. Pontryagin, Topological Groups, 2nd ed., Gordon and Breach, New York, 1966.

[Pr] R. Proctor, Solution of Two Difficult Combinatorial Problems with Linear Algebra, this MONTHLY, 89 (1982) 721–734.

[R] W. Rudin, Principles of Mathematical Analysis, 3rd ed., McGraw-Hill, New York, 1976.

[Se] J. P. Serre, Lie Algebras and Lie Groups, Benjamin, New York, 1965.

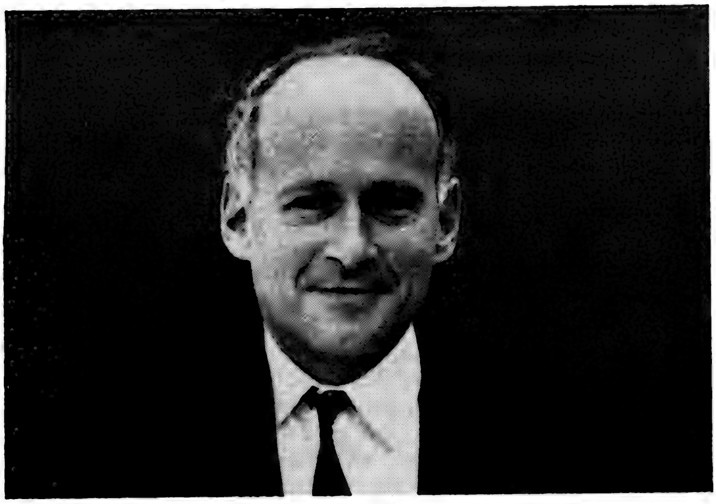
[St] G. Strang, Linear Algebra and its Applications, 2nd ed., Academic Press, New York, 1980.

[Th] W. Thurston, The Topology and Geometry of 3 Manifolds, Princeton University lecture notes.

[V] N. J. Vilenkin, Special Functions and the theory of group representations, A.M.S. Translations of Math. Monographs, no. 22, American Mathematical Society, Providence, R.I., 1968.

[Wg] E. Wigner, Symmetries and Reflections, Indiana University Press, Bloomington, IN, 1967.

[W] H. Weyl, Gruppentheorie und Quantenmechanik, S. Hirzel Verlag, Leipzig, 1928. (English translation: The Theory of Groups and Quantum Mechanics, Dover Publications, New York)



A French mathematician and a Russian one, not necessarily in that order. See p. 647.

holds. Summation extends over all d, $d \| (n, r)$. u is the greatest common square-free unitary divisor of n and r. (A divisor d of c is called unitary if $(d, c/d) = 1$.) $\phi(\mu)$ is the Euler totient [Möbius function].

Solution by Chun-Nip Lee, student, Massachusetts Institute of Technology. The stated formula contains a misprint. A correct formula is

$$\sum \phi(nr/d^2)\, d\mu(d) = \phi(n/u)\phi(r/u)\phi_2(u^2).$$

(*Editor's note.* The greatest common square-free unitary divisor (g.c.s.f.u.d.) of p and p^α, $\alpha \geqslant 2$, is 1; this is not the same as the greatest square free unitary divisor of the greatest common divisor of two numbers.)

Proof by induction on the number of distinct prime factors of (n, r). If $(n, r) = 1$, then each side of the equation equals $\phi(n)\phi(r)$ by multiplicativity of ϕ. Now suppose that the formula is true if (n, r) contains at most k distinct prime factors, $k \geqslant 0$, and that $(n, r) = p^\eta e$, where e contains k distinct prime factors, $\eta \geqslant 1$, and $(p, e) = 1$. Write $n = p^\nu N$ and $r = p^\rho R$, with $(p, N) = (p, R) = 1$. Let U denote the g.c.s.f.u.d. of N and R and u the g.c.s.f.u.d. of n and r.

Let Σ denote the sum on the left side of the formula. We can express Σ as a sum over the divisors of e plus a sum over the remaining divisors of pe, since $\mu(d) = 0$ if d is divisible by the square of a prime. We obtain by multiplicativity

$$\begin{aligned}\sum &= \sum_{d|p^\eta e} \phi(p^{\nu+\rho}NR/d^2)\, d\mu(d)\\ &= \left(\phi(p^{\nu+\rho}) - p\phi(p^{\nu+\rho-2})\right)\sum_{d|e}\phi(NR/d^2)\, d\mu(d).\end{aligned}$$

The induction hypothesis now gives

$$\sum = \left(\phi(p^{\nu+\rho}) - p\phi(p^{\nu+\rho-2})\right)\phi\left(\frac{N}{U}\right)\phi\left(\frac{R}{U}\right)\phi_2(U^2).$$

If $\nu = \rho = 1$, then $u = pU$ and

$$\begin{aligned}\sum &= (p^2 - 2p)\phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(U^2)\\ &= \phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(u^2),\end{aligned}$$

since ϕ_2 is multiplicative and $\phi_2(p^2) = p^2 - 2p$.

If one of ν, ρ exceeds 1, then $u = U$ and

$$\begin{aligned}\sum &= (p^{\nu+\rho} - 2p^{\nu+\rho-1} + p^{\nu+\rho-2})\phi\left(\frac{N}{U}\right)\phi\left(\frac{R}{U}\right)\phi_2(U^2)\\ &= \phi(p^\nu)\phi(p^\rho)\phi\left(\frac{N}{u}\right)\phi\left(\frac{R}{u}\right)\phi_2(u^2)\\ &= \phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(u^2).\end{aligned}$$

This completes the induction.

ANSWERS TO PHOTOS ON PAGE 624

Top: A. G. Kurosh (algebra, and in particular, group theory), taken in 1958. Bottom: L. Schwartz (analysis, and in particular, distributions) taken in 1969.

CENTER SECTION

(Vol. 90, No. 9, November 1983)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S(15-17), P*, L. Judgment Under Uncertainty: Heuristics and Biases. Ed: Daniel Kahneman, Paul Slovic, Amos Tversky. Cambridge U Pr, 1982, xiii + 555 pp, $44.50; $14.95 (P). [ISBN: 0-521-24064-6; 0-521-28414-7] Very interesting collection of articles, dealing primarily with psychological studies of how people estimate probabilities and the types of biases produced by these heuristic methods. A "blemished portrait of human capabilities...emerges from this work," that has serious implications for decision making. Good set of references. RSK

General, S(15-18), P, L*. Winning Ways for Your Mathematical Plays, V. 1: Games in General. Elwyn R. Berlekamp, John H. Conway, Richard K. Guy. Academic Pr, 1982, xxxii + 437 pp, $58.50 (P). [ISBN: 0-12-091101-9] Serious mathematics in whimsical dress: strategic analyses (loaded with non-stop outrageous puns) of several dozen variations on games such as nim, hackenbush, and kayles. The focus of Volume 1 is on the Grundy theory of two-person positional games and their variations (sums, urgent unions, misère) as encoded in Conway's "surreal" numbers (On Numbers and Games, 1976, TR, November 1976). Both terminology (especially puns) and notation are highly specialized (the index lists over 100 symbols that are virtually unique to this book), making casual reading quite difficult. LAS

General, S*, P*, L.** Powers of Ten. Philip and Phylis Morrison. WH Freeman, 1982, 159 pp, $29.95. [ISBN: 0-7167-1409-4] A stunning re-creation on paper of the famous short film "Powers of Ten": each right-hand page contain a view of a picnic in Chicago at different distances, from 10^{25} meters (1 billion light years) to 10^{-16} meters (0.1 fermi), while the left pages offer commentary and other illustrations at each scale. This new still film is framed and enriched by an opening essay on matters of size, concluding commentary, and references. This is the first volume in the new Scientific American Library, a series of books for the general reader. LAS

General, S*, L*. Problem-Solving Through Problems. Loren C. Larson. Prob. Books in Math. Springer-Verlag, 1983, xi + 332 pp, $34. [ISBN: 0-387-90803-X] An anthology of over 700 problems, mostly from the Monthly and Putnam exams, classified by method of solution and embedded in a succession of sections each featuring the power of a single method. Intended especially for students preparing for mathematical competitions, the book is also well suited to self-study by any undergraduate mathematics major who wants to hone the tools of his trade. LAS

Precalculus, T(13: 1), S, L. Precalculus Mathematics, A Functions Approach. Floyd E. Helton, Margaret L. Lial. Scott Foresman, 1983, xi + 514 pp, $19.95. [ISBN: 0-673-15507-2] The central theme is function--definition, graphs, standard precalculus examples: polynomial, log and exponential, trigonometric. Other topics: coordinate geometry, systems of equations, complex numbers, sequences and series. Nice feature: effective use of examples to motivate definitions and development. Some applications. PZ

History, S, L.** James Clerk Maxwell, A Biography. Ivan Tolstoy. U of Chicago Pr, 1981, viii + 184 pp, $6.95 (P). [ISBN: 0-226-80787-8] A short, highly readable account of Maxwell's life--conservative, unassuming, quietly religious--and professional work--daring, revolutionary, powerfully mathematical. Tolstoy, an applied mathematician, conveys clearly both the context and significance of Maxwell's electromagnetic theory without resort to any equations or symbols: he stresses instead the role of analogy in Maxwell's work, illustrated by extensive expository quotations from Maxwell's important papers. LCL

History, S(13-16), L*. A Calculating People: The Spread of Numeracy in Early America. Patricia Cline Cohen. U of Chicago Pr, 1982, x + 271 pp, $22.50. [ISBN: 0-226-11283-7] An intriguing account of the role of arithmetical and statistical reasoning in eighteenth and early nineteenth century America, from publication of Bills of Mortality and impassioned argument (in 1722) about the merits of smallpox innoculation to slavery debates a century later engendered by suspect data on insanity in the 1840 census. Contains many insights into the early American history of statistics

(originally meaning statements, not necessarily quantitative, about affairs of state), including the founding of the American Statistical Society. LAS

Number Theory, P. Multiple Trigonometric Sums. G.I. Arhipov, A.A. Karacuba, V.N. Cubarikov. Proc. of Steklov Inst. of Math., Tom 151. AMS, 1982, viii + 126 pp, $42 (P). [ISBN: 0-8218-3067-8] Theorems about multiple trigonometric sums applied to problems concerning the number of solutions of complicated systems of Diophantine equations. LAS

Algebra, T(18), P. Modules and Rings. F. Kasch, D.A.R. Wallace. London Math. Society Mono., V. 17. Academic Pr, 1982, xiii + 372 pp, $63. [ISBN: 0-12-400350-8] Starting with the fundamental concepts of the theory of rings and modules (early emphasis on the importance of projective and injective modules) the author proceeds to develop a number of important themes that heretofore have not appeared in textbook form (e.g., rings with perfect duality and quasi-Frobenius rings). Numerous exercises. LCL

Finite Mathematics, T(13-14: 1, 2). Applied Finite Mathematics, Third Edition. Howard Anton, Bernard Kolman. Academic Pr, 1982, xi + 593 pp. [ISBN: 0-12-059566-4] Minor modifications, including a few new exercises, and chapter review summaries. Added appendix on algebra review. (First Edition, TR, January 1975; Second Edition, TR, June-July 1978.) LCL

Calculus, S(13). Solutions Manual to Accompany Stein's Calculus and Analytic Geometry, Third Edition. Anthony Barcellos. McGraw-Hill, 1982, $10 (P). [ISBN: 0-07-061155-6] Complete solutions to odd-numbered exercises. LCL

Real Analysis, T*(15: 2, 3), S, L*. Intermediate Real Analysis. Emanuel Fischer. Undergraduate Texts in Math. Springer-Verlag, 1983, xiv + 770 pp, $28. [ISBN: 0-387-90721-1] Assuming (correctly) that students learn little from slick proofs, the author has produced an extraordinarily complete treatment of the first course in real analysis (144 pages until Cauchy sequences). Examples and exercises abound. It would take (at least) a year to cover the material, but the students would know the reasoning as well as the statements of the major ideas in analysis. TAV

Real Analysis, T(17: 2). Real Analysis, Second Edition. Serge Lang. Addison-Wesley, 1983, xiv + 533 pp, $23.95. [ISBN: 0-201-14179-5] In six parts--general topology, Banach spaces and the calculus, functional analysis, integration in measured spaces, integration on locally compact and Euclidean spaces, and global analysis--gives the scope and flavor of this text. Carefully written with extensive problem sets, this contains much more analysis than the title suggests. TAV

Real Analysis, T(15-16: 2). Principles of Real Analysis, Revised Edition. S.C. Malik. Halsted Pr, 1980, viii + 379 pp, $19.95. Typical topics--sets, sequences, functions, integrals, derivatives--clear exposition, a limited number of exercises. The printing is quite poor on grey paper. TAV

Complex Analysis, T(15-16: 1), S, L. Complex Variables with Applications. A. David Wunsch. Addison-Wesley, 1983, viii + 439 pp, $26.95. [ISBN: 0-201-08885-1] A mathematically relatively elementary undergraduate text. Real calculus ideas are reviewed, some proofs simplified, e.g., Cauchy-Goursat via Green's theorem. Notable for the number and depth of physical applications, especially to electricity. List of topics quite standard; order, emphasis, notation less so, consistent with emphasis on applications. E.g., strong sections on residue integrals, integral transforms. Interesting problems, many applied. PZ

Complex Analysis, T*(18: 1, 2), S, P*. Function Theory on Planar Domains: A Second Course in Complex Analysis. Stephen D. Fisher. Wiley, 1983, xiii + 269 pp, $34.95. [ISBN: 0-471-87314-4] The Dirichlet problem, Hardy spaces, inner functions, the corona theorem, and other function-theoretic topics are developed for multiply-connected planar domains, usually with real-analytic boundaries. While there are many references to additional readings, the text is self-contained (at the second-year graduate level). With exercises following each chapter. An excellent text or self-study manual. PZ

Complex Analysis, P. Lecture Notes in Mathematics-978: Quasiconformal Mappings in the Plane: Parametrical Methods. Julian Ławrynoeicz, Jan Krzyz. Springer-Verlag, 1983, vi + 177 pp, $10.50 (P). [ISBN: 0-387-11989-2] From the foreward: "An exposition of analytic properties of quasiconformal mappings in the plane..., a detailed and systematic study of the parametrical method with complete proofs..., and a brief account of variational methods." PZ

Differential Equations, T(17-18: 1, 2), S, P. Shock Waves and Reaction-Diffusion Equations. Joel Smoller. Springer-Verlag, 1983, xxi + 581 pp, $39. [ISBN: 0-387-90752-1] A comprehensive introduction to partial differential equations, emphasizing, but not limited to, the nonlinear topics of the title. Part I, on linear partial differential equations, could serve as an introductory graduate text (but note: no exercise sets). Focus on topological aspects: Part IV treats Morse theory and the Conley index. Most parts should be readable by most mathematicians. PZ

Functional Analysis, S(18), P. Functional Analysis, Holomorphy, and Approximation Theory. Ed: Guido I. Zapata. Lect. Notes in Pure & Appl. Math., V. 83. Dekker, 1983, viii + 458 pp, $55 (P). [ISBN: 0-8247-1634-5] Proceedings of the August 1979 International Seminar on Functional Analysis, Holomorphy, and Approximation Theory at Rio de Janeiro. 19 papers, mostly on functional analysis topics, at the research and advanced expository level. For the nonspecialist: an interesting essay by Dieudonné on the history of functional analysis. PZ

Functional Analysis, T*(17: 1, 2), S, L. First Course in Functional Analysis, Second Edition. Casper Goffman, George Pedrick. Chelsea Pub, 1983, xi + 284 pp, $15.95. [ISBN: 0-8284-0319-8] A second edition (with revised bibliography) of the 1965 first-year graduate text. Unusually well-integrated, concrete point of view, convincingly presented: "Analysis is basic...abstract theories...are primarily of interest as tools...in treating problems in analysis." Many analysis topics motivate and illustrate--Lebesgue theory, reproducing kernels, Haar measure, Fourier series. PZ

Analysis, S(17-18), P. Lecture Notes in Mathematics-977: On Global Univalence Theorems. T. Parthasarathy. Springer-Verlag, 1983, viii + 106 pp, $8.50 (P). [ISBN: 0-387-11988-4] The problem studied is to give conditions for differentiable R^n-valued functions on subsets of R^n to be globally univalent. The first two brief chapters collect rudiments of the theory from real analysis and matrix theory; the remainder of the monograph consists of history, old and new results, applications, and open problems. PZ

Algebraic Geometry, P. The Birational Geometry of Degenerations. Ed: Robert Friedman, David R. Morrison. Progress in Math., V. 29. Birkhauser Boston, 1983, ix + 386 pp, $27.50. [ISBN: 3-7643-3111-9] Nine papers from the Summer Algebraic Geometry Seminar, 1981, at Harvard, studying the geometry of birational equivalence classes of certain completions of families of compact complex manifolds over the punctured disk. The first paper reviews earlier work in the field and surveys the results in the remainder of the volume. PZ

Algebraic Geometry, P. The Curves Seminar at Queen's, Volume II. Ed: Anthony V. Geramita. Papers in Pure & Appl. Math., No. 61. Queen's U, 1982, 216 pp, (P). Four research articles (intersection theory, Hilbert functions) and two expository articles (graded rings, ordinary singularities of curves). LCL

Differential Geometry, P. Nonlinear Analysis on Manifolds. Monge-Ampère Equations. Thierry Aubin. Grund. der math. Wissenschaften, B. 252. Springer-Verlag, 1982, xii + 204 pp, $29.50. [ISBN: 0-387-90704-1] Introduces and treats nonlinear problems on Riemannian manifolds, emphasizing interplay betweeen analysis and geometry. First four chapters are foundational material on analysis of Riemannian manifolds; last four are applications, e.g., to real and complex Monge-Ampère equations. "A reference and...an introduction to research." PZ

Geometry?? Solutions to the Three Historical Problems by Compass and Straightedge. Delvin J. Johnson. Vantage Pr, 1982, v + 30 pp, $6.95. [ISBN: 533-05050-2] A worthless collection of strangely worded "theorems" (e.g., "Given the opportunity, an arc for an acute angle will give the chord for any given angle arc reduced to 1/3.") with three to seven line "proofs." JNC

Geometry, T(14-16: 1), S, L. Foundations of Euclidean and Non-Euclidean Geometry. Richard L. Faber. Pure & Appl. Math., V. 73. Dekker, 1983, xi + 329 pp, $49.75. [ISBN: 0-8247-1748-1] The foundations of Euclidean and non-Euclidean geometry are treated with a strong historical flavor and great pedagogical concern. Half the volume is concerned with a relatively novel detailed development of hyperbolic geometry, its models and its relevance to physics. Many good exercises. SS

Geometry, T(14-16: 1, 2), S, L. Invitation to Geometry. Z.A. Melzak. Wiley, 1983, ix + 225 pp, $29.95. [ISBN: 0-471-09209-6] An unusual and interesting potpourri of geometric topics appropriate for college-level students. Focus is on problems--pure and applied--from Euclidean geometry of two and three dimensions. Not only are most of the problems fascinating, but so are the methods of solution, many of which are developed as general problem-solving techniques for the reader. Rich sets of exercises. SS

Topology, S(15-18), P. Shape Theory: The Inverse System Approach. S. Mardešić, J. Segal. Math. Lib., V. 26. Elsevier North-Holland, 1982, xv + 378 pp, $81.50. [ISBN: 0-444-86286-2] Chapter one presents the shape category of topological spaces with special emphasis on inverse limits and resolutions of spaces. The second chapter treats the algebraic topology of shape theory while chapter three contains a number of surveys of selected areas of shape theory. Many sections of bibliographic notes interspersed throughout the text. There is an appendix on polyhedra and one on Borsuk's approach to shape. Extensive bibliography. List of special symbols. Subject index. RJA

Operations Research, T(17-18: 1), P. Mathematical Aspects of Scheduling and Applications. R. Bellman, A.O. Esogbue, I. Nabeshima. Modern Appl. Math. & Computer Sci., V. 4. Pergamon Pr, 1982, xiv + 329 pp, $35. [ISBN: 0-08-026477-8] A unified treatment of scheduling problems and their solutions. State-of-the-art solution techniques are presented. Some familiarity with dynamic programming and integer programming is assumed. AO

Probability, T(17: 2), P. Approximating Countable Markov Chains. David Freedman. Springer-Verlag, 1983, x + 140 pp, $20. [ISBN: 0-387-90804-8] The third in a "trilogy"--follows the author's Markov Chains and Brownian Motion and Diffusion (TR, June-July 1983). Constructive in approach, the text presumes familiarity with separability, and emphasizes the general (countable) process. A reprint of the 1972 Holden Day edition (TR, January 1973). TAV

Probability, T(18: 2). Stochastic Calculus and Applications. Robert J. Elliott. Applic. of Math., No. 18. Springer-Verlag, 1982, ix + 302 pp, $42. [ISBN: 0-387-90763-7] Presumes probability, measure, and a course in stochastic processes. The author attempts to follow the modern French theory of random processes to a point where applications to system theory and control may be

understood. The development makes use of Martingale methods in stochastic control. Contains an extensive bibliography. TAV

Probability, P. Regular and Stochastic Motion. A.J. Lichtenberg, M.A. Lieberman. Appl. Math. Sci., No. 38. Springer-Verlag, 1983, xxi + 499 pp, $36. [ISBN: 0-387-90707-6] Treats stochastic motion in nonlinear oscillator systems. The main emphasis is on the intrinsic stochasticity of systems when the stochastic motion is generated by the dynamics itself, along with an introduction to chaotic motion in dissipative systems. Contains a comprehensive bibliography. TAV

Probability, T(16: 1), S, L. Studies in the Theory of Random Processes. A.V. Skorokhod. Dover Pub, 1982, viii + 199 pp, $4.50 (P). [ISBN: 0-486-64240-2] In three parts: random processes and stochastic integrals; stochastic differential equations; limit theorems for Markov processes. Differs from the usual treatment in its probabilistic rather than analytic approach. A nice book, an excellent buy. TAV

Probability, T(17). Markov Chains. David Freedman. Springer-Verlag, 1983, xiv + 382 pp, $28. [ISBN: 0-387-90808-0] The first in a "trilogy" on Markov processes, this text goes well beyond the usual treatment. Presumes familiarity with first principles of the theory. Contains numerous examples, no exercises. Both the discrete and continuous chains are investigated. A reprint of the 1971 Holden-Day edition (TR, November 1971). TAV

Probability, P. Probabilistic Metric Spaces. B. Schweizer, A. Sklar. Ser. in Prob. & Appl. Math. Elsevier Sci Pub, 1983, xvi + 275 pp, $39. [ISBN: 0-444-00666-4] By replacing the normal metric on a space with a probability that two points are "close," one can define the PMS of the title. This book traces the development of the theory, various approaches to a triangle inequality and the resulting topological results. Contains a useful and extensive bibliography. TAV

Probability, T(17: 1), P. Statistical Estimation for Stochastic Processes. K. Nanthi. Papers in Pure & Appl. Math., No. 62. Queen's U, 1983, vii + 269 pp, (P). Most texts in stochastic processes avoid the problem of determining the values of the parameters used in the models. This survey fills that gap nicely and could be used as a follow-up to a one semester stochastic processes course. TAV

Probability, P. Application of the Theory of Boundary Value Problems in the Analysis of a Queueing Model with Paired Services. J.P.C. Blanc. Math. Centre Tracts, No. 153. Math Centrum, 1982, 244 pp, Dfl. 33 (P). [ISBN: 90-6196-247-1]

Probability, P. Queues and Point Processes. Peter Franken, et al. Wiley, 1982, 208 pp, $29.95. [ISBN: 0-471-10074-9] Using the recent (1976) concept of embedded marked point processes, the authors attempt to give a unified treatment to several basic problems in queueing systems theory. A variety of generalized arrival processes is considered as well as state-dependent service times. TAV

Statistics, P. Lecture Notes in Statistics-17: Asymptotic Optimal Inference for Non-ergodic Models. Ishwar V. Basawa, David John Scott. Springer-Verlag, 1983, xiii + 170 pp, $15 (P). [ISBN: 0-387-90810-2]

Statistics, T(13-14: 1, 2). A First Course in Business Statistics, Second Edition. James T. McClave, P. George Benson. Dellen Pub, 1983, xvii + 546 pp, $24.95. [ISBN: 0-89517-043-4] Presupposes no college mathematics. The usual topics plus treatments of various probability distributions and a chapter on decision analysis using prior information. Many examples and exercises deal with real data. FLW

Statistics, T*(13: 1). Basic Statistics: An Inferential Approach. Frank H. Dietrich, II, Thomas J. Kearns. Dellen Pub, 1983, xiv + 731 pp, $26.95. [ISBN: 0-89517-044-2] Clearly written elementary text, but with more emphasis on inference-making than most. Topics covered are mostly standard, including regression and correlation, analysis of variance (completely randomized design only), chi-square tests, and non-parametric statistics. Extensive problem sets and sample quizzes (both chapter and cumulative). RSK

Computer Programming, T(13: 1). BASIC Programming, Second Edition. Van Court Hare, Jr. HarBrace J, 1982, xv + 407 pp, $16.95 (P). [ISBN: 0-15-505002-8] The text is divided into three parts: introductory material, including sample start-up procedures; minimal BASIC statements; commonly available extensions of BASIC. Changes from the First Edition (TR, April 1972) include: omission of most of the historical discussion and translations from BASIC to Fortran; replacement of more difficult exercises and examples; new material especially designed for users of personal computers. JRG

Computer Science, S(15-18), P. Lecture Notes in Computer Science-141: GAG: A Practical Compiler Generator. Uwe Kastens, Brigitte Hutt, Erich Zimmermann. Springer-Verlag, 1982, iv + 156 pp, $10 (P). [ISBN: 0-387-11591-9] A GAG is a compiler generator based on attributed grammars. Contains chapters on GAG systems; ALADIN--a language for attributed definitions; an AG for a Pascal-Analyzer; and generating efficient compiler front-ends. Two appendices: one on AG for Pascal and on results of the usage of GAG. References. RJA

Computer Science, S(15-18), P. Lecture Notes in Computer Science-139: An Attribute Grammar for the Semantic Analysis of Ada. J. Uhl, et al. Springer-Verlag, 1982, ix + 511 pp, $22 (P). [ISBN: 0-387-11571-4] An attribute grammar (AG) is a tool used to specify formally the semantics of a

programming language and also a compiler for the language. Divided into three parts: (1) the development of the AG; (2) a rationale for the AG; (3) the AG itself. References. RJA

Computer Science, P. Lecture Notes in Computer Science-145: Theoretical Computer Science. Ed: A.B. Cremers, H.P. Kriegel. Springer-Verlag, 1982, x + 365 pp, $16.50 (P). [ISBN: 0-387-11973-6] Contains the contents of the 3 invited lectures plus the 28 contributed papers presented at the 6th Biannual GI Symposium on Theoretical Computer Science, held in Dortmund, Germany, January 5-7, 1983 and sponsored by the Special Interest Group for Automata and Formal Language Theory of the Gesellschaft für Informatik. RJA

Computer Science, P, L. Deterministic Top-Down and Bottom-Up Parsing: Historical Notes and Bibliographies. Anton Nijholt. Math Centrum, 1983, 118 pp, Dfl. 16,50 (P). [ISBN: 90-6196-245-5] Three bibliographies, each introduced with a brief historical and definitional survey: 406 items on top down parsing of LL grammars (Left to right using Left-most derivations); 561 items on LR parsing (Left to right using Right-most derivations); and 251 items on precedence parsing. LAS

Applications (Artificial Intelligence), S(16-18), P, L. The Handbook of Artificial Intelligence, Volume III. Ed: Paul R. Cohen, Edward A. Feigenbaum. William Kaufmann, 1982, xviii + 639 pp, $45. [ISBN: 0-86576-007-1] This volume includes chapters on models of cognition, automatic deduction, vision, learning and inductive inference, planning and problem solving. Each chapter begins with a subject overview that includes historical and referential information. List of contributors. Bibliography for Volume III is cumulative for three volumes. RJA

Applications (Engineering), T*(15: 1). Reliability Evaluation of Engineering Systems: Concepts and Techniques. Roy Billinton, Ronald N. Allan. Plenum Pr, 1983, x + 349 pp, $42.50. [ISBN: 0-306-41296-9] Presuming no prior experience with probability, the author develops the binomial and Poisson distributions as they apply to reliability, then considers graph theoretic and Markov chain applications. Easily read. An excellent source of examples of an important application of probability. TAV

Applications (Physics), P, L*. Academician Andrei Dmitrievich Sakharov: Collected Scientific Works. Ed: D. ter Haar, D.V. and G.V. Chudnovsky. Dekker, 1982, xvi + 303 pp, $27.50. [ISBN: 0-8247-1714-7] 24 of Sakharov's most important papers from 1947 through 1980, arranged in three groups (plasma physics, cosmology, and field theory), each group introduced by commentary by Sakharov and concluded by commentaries by other physicists. Also includes the text of Sakharov's 1975 Nobel Peace Prize lecture, and lists of Sakharov's "divertissements"--problems in mathematics and physics "undertaken as a domestic pasttime," including Ramsey theory, approximate trisections, number theory algorithms. LAS

Applications (Physics), T(15-16: 1), P, L. A Symmetry Primer for Scientists. Joe Rosen. Wiley, 1983, xiv + 192 pp, $26.95. [ISBN: 0-471-87672-0] First coherent textbook treatment of symmetry and its application to science. Introduction to group theory precedes statement and rigorous derivation of six symmetry principles: the equivalence and symmetry principles for processes in isolated physical systems; the special and general symmetry evaluation principles; the special and general symmetry evolution principles. Some examples and problems assume understanding of quantum theory. JRG

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Eastern Pennsylvania and Delaware Section

The spring meeting of the Eastern Pennsylvania and Delaware Section was held at Millersville State College, Millersville, Pennsylvania on April 9, 1983. It was devoted to "Mathematics and Artificial Intelligence." About 87 persons attended.

programming language and also a compiler for the language. Divided into three parts: (1) the development of the AG; (2) a rationale for the AG; (3) the AG itself. References. RJA

Computer Science, P. Lecture Notes in Computer Science-145: Theoretical Computer Science. Ed: A.B. Cremers, H.P. Kriegel. Springer-Verlag, 1982, x + 365 pp, $16.50 (P). [ISBN: 0-387-11973-6] Contains the contents of the 3 invited lectures plus the 28 contributed papers presented at the 6th Biannual GI Symposium on Theoretical Computer Science, held in Dortmund, Germany, January 5-7, 1983 and sponsored by the Special Interest Group for Automata and Formal Language Theory of the Gesellschaft für Informatik. RJA

Computer Science, P, L. Deterministic Top-Down and Bottom-Up Parsing: Historical Notes and Bibliographies. Anton Nijholt. Math Centrum, 1983, 118 pp, Dfl. 16,50 (P). [ISBN: 90-6196-245-5] Three bibliographies, each introduced with a brief historical and definitional survey: 406 items on top down parsing of LL grammars (Left to right using Left-most derivations); 561 items on LR parsing (Left to right using Right-most derivations); and 251 items on precedence parsing. LAS

Applications (Artificial Intelligence), S(16-18), P, L. The Handbook of Artificial Intelligence, Volume III. Ed: Paul R. Cohen, Edward A. Feigenbaum. William Kaufmann, 1982, xviii + 639 pp, $45. [ISBN: 0-86576-007-1] This volume includes chapters on models of cognition, automatic deduction, vision, learning and inductive inference, planning and problem solving. Each chapter begins with a subject overview that includes historical and referential information. List of contributors. Bibliography for Volume III is cumulative for three volumes. RJA

Applications (Engineering), T*(15: 1). Reliability Evaluation of Engineering Systems: Concepts and Techniques. Roy Billinton, Ronald N. Allan. Plenum Pr, 1983, x + 349 pp, $42.50. [ISBN: 0-306-41296-9] Presuming no prior experience with probability, the author develops the binomial and Poisson distributions as they apply to reliability, then considers graph theoretic and Markov chain applications. Easily read. An excellent source of examples of an important application of probability. TAV

Applications (Physics), P, L*. Academician Andrei Dmitrievich Sakharov: Collected Scientific Works. Ed: D. ter Haar, D.V. and G.V. Chudnovsky. Dekker, 1982, xvi + 303 pp, $27.50. [ISBN: 0-8247-1714-7] 24 of Sakharov's most important papers from 1947 through 1980, arranged in three groups (plasma physics, cosmology, and field theory), each group introduced by commentary by Sakharov and concluded by commentaries by other physicists. Also includes the text of Sakharov's 1975 Nobel Peace Prize lecture, and lists of Sakharov's "divertissements"--problems in mathematics and physics "undertaken as a domestic pasttime," including Ramsey theory, approximate trisections, number theory algorithms. LAS

Applications (Physics), T(15-16: 1), P, L. A Symmetry Primer for Scientists. Joe Rosen. Wiley, 1983, xiv + 192 pp, $26.95. [ISBN: 0-471-87672-0] First coherent textbook treatment of symmetry and its application to science. Introduction to group theory precedes statement and rigorous derivation of six symmetry principles: the equivalence and symmetry principles for processes in isolated physical systems; the special and general symmetry evaluation principles; the special and general symmetry evolution principles. Some examples and problems assume understanding of quantum theory. JRG

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Eastern Pennsylvania and Delaware Section

The spring meeting of the Eastern Pennsylvania and Delaware Section was held at Millersville State College, Millersville, Pennsylvania on April 9, 1983. It was devoted to "Mathematics and Artificial Intelligence." About 87 persons attended.

Invited Addresses:

"Grammatical Representations in Natural Language Processing: Some Mathematical Results," by Aravind K. Joshi, University of Pennsylvania.
"Artificial Intelligence: Basic Principles and Philosophical Issues," by Jaime G. Carbonell, Carnegie-Mellon University.
"Logic and Its Uses in Artificial Intelligence and Databases," by Jack Minker, University of Maryland.
"Expert Systems," by Casimir A. Kulikowski, Rutgers University.

Wisconsin Section

The fifty-first annual meeting of the Wisconsin Section met April 15-16, 1983 at the University of Wisconsin Center, Washington County, West Bend, Wisconsin.

Invited Addresses:

"What To Do Till the Computer Scientist Goes Away," by Paul Halmos, Indiana University.
"Topological, Combinatorial and Geometric Fractals," by James Cannon, University of Wisconsin at Madison.
"Products of Involutions: Is 2 x 2 Always 4?" by Paul Halmos, Indiana University.

Short Presentations:

"An Anthology of Non-Measurable Sets," by Jonathan Lewin, University of Wisconsin at Oshkosh.
"Computer Languages for Distributive Processing," by David Hayes (Student, Ripon College).
"A Case Study in the Development of Microcomputer Software for Mathematics," by Ed Gade, University of Wisconson at Oshkosh.
"Independence of Axioms," by John Oman, University of Wisconsin at Oshkosh.
"Solving Max-Min Problems Using Inequalities," by Robert Horton, University of Wisconsin at Whitewater.
"Inverted Plane Curves," by Orville Bierman, University of Wisconsin at Eau Claire.
"An Introduction to Secret Codes," by Linda Deneen, Beloit College.
"Probabilistic Arguments in Queueing Models and Their Regulation," by Nasser Hadidi, University of Wisconsin at Stout.
"On Inequalities," by Steve Krevisky, University of Wisconsin Center, Washington County.
"Constructing Projective Modules," by Andrew Machett, University of Wisconsin at LaCrosse.
"Analytic Function Theory for Operators," by James Walker, University of Wisconsin at Eau Claire.
"An Empirical Laboratory Course to Accompany Calculus I," by Dan Kalman, University of Wisconsin at Green Bay.

Panel Discussion:

"Mathematicians Teaching Computer Science," by Lois Brualdi, Edgewood College; Tom Napps, Lawrence University; Steve Senger, University of Wisconsin at LaCrosse; Bhagat Singh, University of Wisconsin Center at Manitowoc; Larry Thorsen, St. Norbert College.

Northeastern Section

The Spring meeting of the Northeastern Section was held at Bowdoin College in Brunswick, Maine on June 17-18, 1983. Approximately 90 participants attended the meeting.

Invited Addresses:

"Philosophers Who Think About Themselves: In Defense of Recreational Mathematics," by Raymond M. Smullyan, Lehman College, CUNY, Indiana University.
"New Directions--New Challenges in College Mathematics," by Donald L. Kreider, Dartmouth College.

Short Presentations:

* "Designing an Introductory Course in Applied Mathematics," by Igor Najfeld, Tufts University.
* "Mathematical Modeling: The Missing Component," by William E. Boyce, Rensselaer Polytechnic Institute.
* "How Are We Doing With the Next Generation?," by Florence D. Jacobson, Director of the Women and Mathematics Program for Connecticut, and Professor Emerita, Albertus Magnus College.

Computer Use Workshops:

* "BASIC for Beginners," by Patricia Strauss, Rhode Island College.
* "PASCAL for the Non-Believer," by James W. Uebelacker, University of New Haven.
* "MAT LAB," by I. Gary Rosen, Bowdoin College.
* "muMATH," by David Ryan, Ansonia High School and University of Hartford.
* "SCRIPSIT," by Raymond J. McGivney, University of Hartford, and Edwin A. Rosenberg, Western Connecticut State University.

Contributed Papers:

"Computer-Generated Stereographic Images for Calculus of Several Variables," by Kenneth G. Hamilton, Colby College.
"Fear of Calculus," by Stephanie F. Troyer, University of Hartford.
"Calculus and Computer Modeling: Not Calculus as Usual," by Kenneth Hoffman, Hampshire College.

Metropolitan New York Section

The forty-second annual meeting of the Metropolitan New York section was held at St. John's University on May 7, 1983. Approximately 150 persons were in attendance.

Invited Presentations:

"Iteration of Complex Quadratic Polynomials," by Dennis Sullivan, Einstein Professor of Mathematics, Queens College.
"Alternatives for the Freshman Mathematics Curriculum: Discrete Mathematics vs. the Calculus," by Ronald Douglas, SUNY at Stony Brook, and Anthony Ralston, SUNY at Buffalo.

Panel Discussion:

"Meeting the Needs of Mathematical Talented High School Students," by Robert Cowen (Moderator), Queens College. Panelists: Carl Goodman, Cardozo High School; Irwin Kaufman, New York City Board of Education; David Kelly, Hampshire College; Edwin Moise, Queens College; and Mark Saul, Bronx High School of Science.

Short Presentations:

"Microcomputers and Student Questionnaires in a Beginning Course in Statistics," by Geoffrey Akst, Manhattan Community College.
"The Arkin-Hoggatt Game," by J. Arkin, Spring Valley, New York, and V.E. Hoggatt, San Jose, California.
"A Comparison of the Distribution of Women's and Men's Calculus Grades," by Sherry Blackman, The College of Staten Island.
"Convergence and Divergence of Infinite Series--A Probability Approach Through the St. Petersburg Paradox," by Allan J. Ceasar, U.S. Merchant Marine Academy.
"Hamilton's Number Couples," by Alan Chutsky, Queensborough Community College.
"Statistics through the Eye of the Computer," by Florence S. Gordon, Adelphi University.
"Solved and Unsolved Problems in Axiomatic Geometry," by Martha Harrell, St. John's University.
"Real Variable Inversion of Laplace Transforms," by Harvey J. Hinden, Hi-Tech Editorial, Inc.
"Quantitative Modelling Techniques for Crime Prediction," by Alan Hoenig, John Jay College of Criminal Justice.
"On the Distribution of Planar and Parabolic Points on Developable Surfaces," by Martin Lewinter, SUNY at Purchase.
"A Matrix Equation for the Group Inverse," by Maurice Machover, St. John's University.
"The Right Student in the Right Mathematics Course: Placement with the CUNY Mathematics Test or An Instructor-Developed Test," by Seymour W. Pustilnik, New York City Technical College.
"A Classical Tensor Theory of Gravitation which Arises from the Principle of Equivalence," by David Shelupsky, The City College.
"Math Enrichment for Elementary School Girls, A Pilot Study," by Helga Schwartz, Queensborough Community College.
"Using Microcomputers to Improve Statistics Teaching," by Lawrence Sher, Borough of Manhattan Community College.
"A Poisson Process with a Markovian Intensity Function," by Frederick Solomon, SUNY at Purchase.
"A Doubly Stochastic Poisson Process with Intensity Function Generated by Birth and Death Process," by Frederick Solomon, SUNY at Purchase.

Awards:

Noam D. Elkies of Columbia University received the Section award for the highest regional score in the William Putnam Mathematical Competition.

Douglas Jongreis of G.W. Hewlett High School, Michael Reid of Brooklyn Technical High School, and David Zuckerman of Stuyvesant High School received the Charles Salkind Awards. These three students tied for the highest regional score in the MAA High School Mathematics Contest.

Texas Section

The annual meeting of the Texas section was held at North Texas State University in Denton, Texas on April 8-9, 1983. Registered meeting attendance was 266.

Invited Lectures:

"Paradoxical Coverings of the Real Line," by Ivan Niven, University of Oregon and President of MAA.
"Duality in Mathematics (The Role of Compact Hausdorff Spaces in Mathematics)," by S. Kakutani, Yale University.
"Perceptions Regarding Discrete Mathematics and Calculus for the Freshman Year," by R.D. Anderson, Louisiana State University.
"The State of School Mathematics in Texas," by Barbara Montalto, Texas Education Agency.

Contributed Papers:

"Coloring Graphs and Combinatorial Geometries," by Joseph Kung, North Texas State University.
"A Note on $D_x(x^r)$ for Rational r," by Peter Lindstrom, North Lake College.
"Some Theory and Application of Geometric Inequalities," by J.M. Stark, Lamar University.
"Some Mappings of Sequence Spaces," by David F. Dawson, North Texas State University.
"Commutativity and Norm Conditions in Banach Algebras," by Victor A. Belfi, Texas Christian University.
"Approximating a Function by Means of an M-polynomial," by Frank N. Huggins, University of Texas at Arlington.
"S-essentially T_1 Spaces," by Charles Dorsett, Louisiana Tech University.
"Local Weights in Metric Spaces," by Jim Bradford, Abilene Christian University.
"Approximating the Length of Material on a Cylindrical Roll," by Frank H. Mathis and Danny W. Turner, Baylor University.
"Symmetric Positive Systems with Nonlinear Boundary Conditions," by Alfonso Castro, Southwest Texas State University.
"Expanding Mappings and Fixed Points," by A.A. Gillespie and B.B. Williams, University of Texas at Arlington.
"Differential Equations with Piecewise Constant Delays," by Joseph Wiener, Pan American University.
"An Application of Discrete Cross-correlation," by Philip R. Sanfilippo, Farmers Branch, Texas.
"Cloud and Precipitation Analysis," by C. Bandy and R. Torrejon, Southwest Texas State University.
"TeX at Texas A&M University," by Norman W. Naugle, Texas A&M University.
"Scientific Output Problems in Microcomputers," by Roy Dean Alston, Stephen F. Austin State University.
"Obtaining Sommerfeld's Diffraction Integral via Fourier Transform," by E. Dennis Huthnance, Midwestern State University.
"Some Recent Experiences with Logo," by Norman W. Naugle, Texas A&M University.
"Computer Literacy: A Call for the Problem-solving Approach," by Michael Murphy and Nancy Rich, University of Houston-Downtown College.
"A Proof-teaching Computer System," by Michael K. Jones, Fort Worth, Texas.
"Teaching Mathematics in Africa," by Neal Hart, Sam Houston State University.
"Secondary Mathematics Dilemma: An Elementary Solution," by Bill Aslan, East Texas State University.
"Another Approach in Teaching Arithmetic and Algebraic Fractions," by Bella Wiener, Pan American University.
"The Fascinating Fur Farmer," by G. Marvin Eargle, Appalachian State University, Boone, North Carolina.
"Injecting Vigor as a Prelude to Rigor or How to Teach College Algebra Like a Real Math Course," by G. Edgar Parker, Pan American University.
"Preradicals Induced by Homomorphisms," by Ed Oxford, Baylor University.
"The Splitting of the Kunneth Sequence," by Andre Deutz, Texas A&M University at Galveston.
"Solving Radical Equations," by John Huber, Pan American University.
"Axioms for C*-Algebras: A Survey from 1943-1983," by Robert S. Doran, Texas Christian University.
"The Dunford-Pettis Property," by Tommy Leavelle, North Texas State University.
"The Existence of Nonzero Maps from Injectives to Projectives," by Jeffrey W. Neslen, Texas Christian University.
"On G-finite Algebras," by Joseph M. Szucs, Texas A&M University at Galveston.

Student Papers:

"A Linear Analysis of Liniger and Ruegsegger's Mathematical Model of Fibrinolysis," by Joseph F. Gerda, Jr., University of Texas at Arlington.
"Approximation of e^x Using Maclaurin Series and Range Reduction," by Robert F. Jones, Sam Houston State University.
"Oscillation Properties of Solutions of First Order Linear Difference Equations," by Errol Moncriffe, Texas Southern University.
"Schensted's Algorithm and Young Tableaux," by David C. Sutherland, North Texas State University.
"On the Domain of Existence of First Order Partial Differential Equations," by M.E. Brewster and R. Kannan, University of Texas at Arlington.
"Group Representation Theory with Applications," by Brice McIntyre, East Texas State University.
"Unconditional and Absolute Convergence," by Catherine Abbott, North Texas State University.

A "premeeting" Developmental Mathematics Workshop was sponsored by Addison-Wesley Publishing Company. The speaker for this session was F. Demana of Ohio State University who spoke on "Articulation Between High Schools and Colleges."

WHEN IS $\mathbb{R}^2$ A DIVISION ALGEBRA?

STEVEN C. ALTHOEN AND LAWRENCE D. KUGLER

Department of Mathematics, The University of Michigan-Flint, Flint, MI 48503

Under usual complex multiplication, the vector space $\mathbb{R}^2$ is a division algebra, which is to say for all u, v, and w

(i) $u \cdot (v + w) = u \cdot v + u \cdot w$ and $(u + v) \cdot w = u \cdot w + v \cdot w$;

(ii) $a(u \cdot v) = (au) \cdot v = u \cdot (av)$ for every real number a (making $\mathbb{R}^2$ an algebra); and

(iii) for every v and $u \neq 0$ in $\mathbb{R}^2$, the equations $u \cdot x = v$ and $y \cdot u = v$ have unique solutions in $\mathbb{R}^2$; (making $\mathbb{R}^2$ a division algebra).

Condition (iii) is equivalent to

(iii′)* $$u \cdot v = 0 \text{ implies } u = 0 \text{ or } v = 0.$$

The equivalence of conditions (iii) and (iii′) follows from the invertibility of left and right translation (both linear transformations) by nonzero elements of the algebra.

Are there other ways of defining multiplication to make $\mathbb{R}^2$ a division algebra? There are, of course, division algebras of higher dimension, for example, the quaternions and Cayley numbers. It is well known that multiplication in these systems is completely determined by a multiplication table for a basis and that both are noncommutative, while the Cayley numbers are nonassociative.

These facts suggest two things about an investigation of different multiplications for $\mathbb{R}^2$. First, it suffices to study basis multiplication since the multiplicative structure of an algebra is completely determined by a multiplication table for any basis. The proof of this statement involves a straightforward application of properties (i) and (ii) above. Second, the investigation should include algebras that are noncommutative, nonassociative, or even lack a unit. Indeed, the following theorem shows that without easing conditions on the multiplication the search ends abruptly.

THEOREM 1. *If a 2-dimensional real division algebra either has a unit or is associative, it is $\mathbb{C}$, the complex numbers.*

Proof. Suppose the algebra has a unit, e. Extend $\{e\}$ to a basis $\{e, u\}$ for $\mathbb{R}^2$. Let $v = u^2$. The multiplication table

$$\begin{array}{c|cc} \cdot & e & u \\ \hline e & e & u \\ u & u & v \end{array}$$

determines the algebra. It is obvious that multiplication is commutative on the basis and it is not hard to check that it is associative there as well. It is also easy to verify that if the basis elements of a finite dimensional algebra commute (associate), then the algebra is commutative (associative). By a famous theorem of Frobenius (see [3], [6]) this algebra is $\mathbb{C}$.

S. C. Althoen: I received my Ph.D. in algebraic topology from the City University of New York. I then taught two years at Hofstra University. For the last seven years I have been at The University of Michigan-Flint. My nonmathematical interests include running and genealogy. Of the many individuals who have influenced my mathematical development, I wish to mention Robert M. McLeod at Kenyon College and my thesis advisor, Eldon Dyer.

L. D. Kugler: I received my doctorate from the University of California, Los Angeles, where I was a student of Abraham Robinson, working in nonstandard analysis. I have been at The University of Michigan-Flint since 1966, where I now am Dean of the College of Arts and Sciences. To maintain sanity, I continue to teach and also play the bass clarinet.

*Some authors replace (iii′) with the existence of multiplicative inverses. This is equivalent only in the associative, finite-dimensional case (see [1]). Also, solvability of just one equation in (iii) is, in fact, equivalent to (iii′).

Suppose the algebra is associative. Since it is a division algebra, its nonzero elements form a semigroup for which all equations $u \cdot x = v$ and $y \cdot u = v$ have unique solutions. Such a semigroup must be a group [2], so that the algebra has a unit and must be $\mathbb{C}$ as before. ■

The associativity assumption in Theorem 1 can be weakened. For example, the same result holds for *alternative algebras* in which

$$u \cdot (u \cdot v) = (u \cdot u) \cdot v \quad \text{and} \quad u \cdot (v \cdot v) = (u \cdot v) \cdot v$$

for all u and v. To see this, investigate the product $(u + v) \cdot [(u + v) \cdot u]$.

Commutativity alone does not force a 2-dimensional division algebra to be $\mathbb{C}$. Consider the multiplication table

$\cdot$	u	v
u	$2u$	v
v	v	$-u$

This defines a division algebra, for if $a = a_1u + a_2v \neq 0$, $b = b_1u + b_2v$, and $x = x_1u + x_2v$, then the equation

$$a \cdot x = b$$

yields the system

$$\begin{cases} 2a_1x_1 - a_2x_2 = b_1 \\ a_2x_1 + a_1x_2 = b_2 \end{cases}$$

which has a unique solution for any vector b since the determinant $2a_1^2 + a_2^2$ is nonzero. This algebra is not associative: $(u \cdot u) \cdot v = 2v$ while $u \cdot (u \cdot v) = v$; hence it cannot be $\mathbb{C}$.

Now that we have an example of a 2-dimensional division algebra that is not $\mathbb{C}$, it is natural to ask: Can we find and classify all 2-dimensional real division algebras?

1. Idempotents. We adopt the standard approach to the classification problem and consider the general multiplication table for a 2-dimensional division algebra

TABLE 1

$\cdot$	u	v
u	$a_{11}u + b_{11}v$	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	$a_{22}u + b_{22}v$

where $\{u, v\}$ is a basis and a_{ij}, b_{ij} are real numbers. Classification will be easier if we can simplify this table. Complex multiplication suggests the possible simplification $a_{11} = 1$, $b_{11} = 0$. Thus, we are led to consider *idempotents*: nonzero elements w for which $w^2 = w$.

Conditions for an idempotent are easily obtained. Set $w = xu + yv$, x, y in $\mathbb{R}$.

Since $\{u, v\}$ is a basis, the equation $w^2 = w$ yields

$$\begin{cases} x = a_{11}x^2 + (a_{12} + a_{21})xy + a_{22}y^2 \\ y = b_{11}x^2 + (b_{12} + b_{21})xy + b_{22}y^2. \end{cases} \tag{1}$$

Nontrivial solutions of these conic equations correspond to idempotents. A conic is determined by five points no three of which are collinear. Thus, one expects these conics to intersect at 0, 1, 2, 3, or infinitely many points other than the origin, yielding algebras with the corresponding number of idempotents. Each of these cases actually occurs. Examples $i = 0$, 1, 2, and 3 are algebras with exactly i idempotents. Example 4 is an algebra with infinitely many idempotents.

EXAMPLE 0. The trivial algebra: all products zero.

EXAMPLE 1. $\mathbb{C}$.

EXAMPLE 2.

$\cdot$	u	v
u	u	$\frac{1}{2}u+v$
v	$\frac{1}{2}u+v$	v

EXAMPLE 3.

$\cdot$	u	v
u	$11u+6v$	$4u+7v$
v	$6u+5v$	$2u+5v$

The idempotents are: $2u-3v$, $-u+2v$, and $(1/23)(u+v)$.

EXAMPLE 4.

$\cdot$	u	v
u	u	$\frac{1}{2}u+\frac{1}{2}v$
v	$\frac{1}{2}u+\frac{1}{2}v$	v

Here $tu+(1-t)v$ is idempotent for all real numbers t.

If we now restrict our attention to *division* algebras, how many idempotents can there be?

THEOREM 2. *Every 2-dimensional real division algebra has at least one idempotent.*

Proof. In Table 1, if $b_{11}=0$, then $a_{11}\neq 0$, for otherwise $u^2=0$, contradicting condition (iii$'$). Then $(1/a_{11})u$ is idempotent. Similarly, if $a_{22}=0$, then $(1/b_{22})v$ is idempotent. These idempotents correspond to an intersection of the conic equations (1) at a point on either the x or y axis. If $b_{11}\neq 0$ and $a_{22}\neq 0$, then perhaps a rotation would yield a new basis $\{u', v'\}$ such that $(u')^2=fu'$ from which we would get the idempotent $(1/f)u'$. To this end, let

$$u'=cu+sv$$
$$v'=-su+cv,$$

where c and s are the cosine and sine of the rotation angle. Then

$$u'\cdot u'=f(c,s)u'+g(c,s)v', \tag{2}$$

where f and g are the homogeneous cubic polynomials

$$f(x,y)=a_{11}x^3+(a_{12}+a_{21}+b_{11})x^2y+(b_{12}+b_{21}+a_{22})xy^2+b_{22}y^3,$$
$$g(x,y)=b_{11}x^3+(b_{12}+b_{21}-a_{11})x^2y-(a_{12}+a_{21}-b_{22})xy^2-a_{22}y^3.$$

Let r be a real root of the cubic equation $g(1,y)=0$ and set

$$c=1/\sqrt{r^2+1},\quad s=r/\sqrt{r^2+1}.$$

Since g is homogeneous, $g(c,s)=0$. It follows from (2) that

$$u'\cdot u'=f(c,s)u',$$

where $f(c,s)\neq 0$ because $(u')^2\neq 0$. Thus, $u'/f(c,s)$ is an idempotent. ∎

COROLLARY. *Every 2-dimensional real division algebra D contains a subalgebra isomorphic to $\mathbb{R}$.*

Proof. The injection i: $\mathbb{R}\to D$ given by $i(r)=r\cdot u$ is an isomorphism for any idempotent u. ∎

Theorem 2 implies that every 2-dimensional real division algebra has a multiplication table of the form

TABLE 2

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	$a_{22}u + b_{22}v$

In what follows we assume that the multiplication table is in this form.

The table of Example 4 is already in the form of Table 2. However, this algebra with infinitely many idempotents is not a division algebra since $(u - v)^2 = 0$.

This example points out the need for an effective way to determine whether an algebra is a division algebra. Paralleling our earlier example, let

$$a = a_1u + a_2v \neq 0, b = b_1u + b_2v, x = x_1u + x_2v.$$

Then by Cramer's Rule, the equation

$$a \cdot x = b$$

always has a solution if and only if

$$Q(a_1, a_2) = \begin{vmatrix} a_1 + a_2a_{21} & a_1a_{12} + a_2a_{22} \\ a_2b_{21} & a_1b_{12} + a_2b_{22} \end{vmatrix} \neq 0$$

for all pairs of real numbers $(a_1, a_2) \neq (0,0)$. If we set

$$A_1 = \begin{vmatrix} a_{12} & b_{12} \\ a_{21} & b_{21} \end{vmatrix}, \quad A_2 = \begin{vmatrix} a_{21} & b_{21} \\ a_{22} & b_{22} \end{vmatrix},$$

the quadratic form Q has discriminant

$$(b_{22} - A_1)^2 - 4b_{12}A_2.$$

Therefore, $Q(x, y)$ is nonzero for all $(x, y) \neq (0,0)$ if and only if this discriminant is negative.

THEOREM 3. *An algebra described by Table 2 is a division algebra if and only if*

$$(b_{22} - A_1)^2 < 4b_{12}A_2, \tag{3}$$

where

$$A_1 = \begin{vmatrix} a_{12} & b_{12} \\ a_{21} & b_{21} \end{vmatrix} \quad A_2 = \begin{vmatrix} a_{21} & b_{21} \\ a_{22} & b_{22} \end{vmatrix}. \blacksquare$$

An example of a 2-dimensional division algebra with only one idempotent is $\mathbb{C}$. Can a 2-dimensional division algebra have two or more idempotents?

If the algebra has two distinct idempotents u and v, then $\{u, v\}$ forms a basis. Otherwise, $u = av, a \neq 0$ and

$$av = u = u^2 = (av)^2 = a^2v^2 = a^2v$$

so that $a = a^2$, $a = 1$ and $u = v$. Table 1 now takes the form

TABLE 3

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	v

LEMMA. *In a 2-dimensional real division algebra with multiplication table given by Table 3,* $(a_{12} + a_{21})(b_{12} + b_{21}) \neq 1$.

Proof. Let $A = a_{12} + a_{21}$, $B = b_{12} + b_{21}$, $C = a_{12}b_{21}$, $D = a_{21}b_{12}$. If $AB = 1$, then

$$C + D + a_{12}b_{12} + a_{21}b_{21} = 1. \tag{4}$$

The discriminant condition (3) becomes

$$C^2 + D^2 - 2CD - 2C - 2D + 1 < 0. \tag{5}$$

From (5) we see that neither C nor D is zero. Thus

$$b_{21} = C/a_{12} \neq 0 \quad \text{and} \quad b_{12} = D/a_{21} \neq 0.$$

If we substitute these expressions into (4) and let $E = a_{12}/a_{21}$, we obtain

$$DE^2 + (C + D - 1)E + C = 0.$$

Thus E is a solution to a quadratic; so the discriminant

$$C^2 + D^2 - 2CD - 2C - 2D + 1 \geqslant 0. \tag{6}$$

Since (5) and (6) are contradictory, the lemma is proved. ∎

THEOREM 4. *A 2-dimensional real division algebra has exactly one, two, or three idempotents.*

Proof. If the algebra has 2 idempotents u and v, then it can be presented as in Table 3. For such a table, the system (1) of conics reduces to

$$\begin{cases} x = x^2 + Axy \\ y = Bxy + y^2 \end{cases}$$

so that

$$\begin{cases} x = 0, & x = -Ay + 1 \\ y = 0, & y = -Bx + 1. \end{cases}$$

If $A = 1$ or $B = 1$, then since $AB \neq 1$, these degenerate conics intersect at precisely three points; $(0,0)$, $(1,0)$, and $(0,1)$, yielding exactly two idempotents u and v. Now suppose $A \neq 1$ and $B \neq 1$. The lemma implies that the lines $x = -Ay + 1$ and $y = -Bx + 1$ are neither parallel nor coincident. Thus, they will intersect to determine the unique third idempotent

$$\frac{A-1}{AB-1}u + \frac{B-1}{AB-1}v. \blacksquare$$

How can algebras with two or three idempotents be distinguished from those with only one? The function $g(1, y)$ in the proof of Theorem 2 does this. First, note that $g(1, y)$ is defined even if $b_{11} = 0$. Now for algebras with Table 2 we have

$$g(1, y) = -y\left[a_{22}y^2 + (A - b_{22})y + (1 - B)\right].$$

Distinct real roots of this polynomial yield distinct cosines, c, which in turn produce distinct idempotents u'. The root 0 corresponds to the idempotent u. The quadratic factor of $g(1, y)$ will have no real roots if and only if its discriminant is negative. Thus, we have proved the following proposition.

PROPOSITION. *An algebra has exactly one idempotent if and only if*

$$(A - b_{22})^2 < 4a_{22}(1 - B), \tag{7}$$

where $A = a_{12} + a_{21}$ *and* $B = b_{12} + b_{21}$. ∎

2. The Two-Idempotent Case. An algebra with exactly two idempotents u and v can be presented by Table 3 where either $A = 1$ or $B = 1$ but not both. Interchanging basis elements switches A and B, so we may take $A = 1$ and $B \neq 1$. Since automorphisms preserve idempotents, a two-idempotent algebra is uniquely determined by a table with $A = 1$.

THEOREM 5. *Let $\mathcal{A}$ and $\mathcal{A}'$ be 2-dimensional real division algebras, each with exactly two idempotents. Then $\mathcal{A}$ and $\mathcal{A}'$ are isomorphic if and only if their presentations by Table 3 with $A = 1$ are identical.* ■

Example 2 above defines a 2-dimensional real division algebra with exactly two idempotents.

3. The Three-Idempotent Case. For an algebra with three idempotents u, v, and w, any pair, say $\{u, v\}$, may be used as a basis for a multiplication table:

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	v

The third idempotent, w, was given in the proof of Theorem 4.

It is straightforward to construct the multiplication tables based on $\{u, w\}$ and $\{v, w\}$:

$\cdot$	u	w
u	u	$c_{12}u + b_{12}w$
w	$c_{21}u + b_{21}w$	w

$\cdot$	v	w
v	v	$c_{12}v + a_{21}w$
w	$c_{21}v + a_{12}w$	w

where

$$c_{12} = \frac{a_{12}b_{21} - (b_{12} - 1)(a_{21} - 1)}{AB - 1}$$

and c_{21} is obtained by interchanging subscripts. The classification theorem for three-idempotent algebras now follows from the fact that an algebra isomorphism preserves idempotents and is completely determined by its effect on a basis.

THEOREM 6. *Let $\mathcal{A}$ and $\mathcal{A}'$ be 2-dimensional real division algebras, each with three idempotents. Then $\mathcal{A}$ and $\mathcal{A}'$ are isomorphic if and only if there is a correspondence between the idempotents for which the appropriate multiplication tables are identical.* ■

4. The Single-Idempotent Case. For a single-idempotent division algebra, the coefficient a_{22} in Table 2 must be nonzero, for otherwise $(1/b_{22})v$ is another idempotent. One approach to simplification is a basis change to make b_{22} vanish. In fact, the next theorem shows when it is possible to simplify Table 2 to

TABLE 4

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	$\pm u$

THEOREM 7. *Let $\mathcal{A}$ be a single-idempotent 2-dimensional real division algebra given by Table 2. If $b_{12} + b_{21} = B \neq 0$, then $\mathcal{A}$ has a basis $\{u, w\}$ for which $w \cdot w = \pm u$. Furthermore, w is unique up to sign.*

Proof. We must find $\alpha, \beta \neq 0$ so that $w = \alpha u + \beta v$ satisfies $w \cdot w = \pm u$. Calculating $w \cdot w$ by Table 2 and equating coefficients yields

$$\begin{cases} \alpha^2 + \alpha\beta A + \beta^2 a_{22} = \pm 1 \\ \alpha\beta B + \beta^2 b_{22} = 0 \end{cases},$$

where $A = a_{12} + a_{21}$.

If $b_{22} = 0$, then since $B \neq 0$, we have $\alpha = 0$ and $\beta = \pm 1/\sqrt{|a_{22}|}$ so that $w = \pm v/\sqrt{|a_{22}|}$.

If $b_{22} \neq 0$, we obtain

$$\alpha^2(1 - AB/b_{22} + B^2 a_{22}/b_{22}^2) = \pm 1,$$

which can be solved uniquely up to sign for α by selecting the sign on the right to match the sign of the coefficient of α^2. ■

Here is an example:

EXAMPLE 5.

$\cdot$	u	v
u	u	v
v	$u+v$	$-u$

The apparent table dependence of the condition $B \neq 0$ seems to complicate the classification problem. But a straightforward investigation of the basis change from $\{u, v\}$ to $\{u, w = \alpha u + \beta v\}$, $\beta \neq 0$, using Table 2 shows that

$$u \cdot w = (\alpha + a_{12}\beta - b_{12}\alpha)u + b_{12}w, \tag{8}$$

$$w \cdot u = (\alpha + a_{21}\beta - b_{21}\alpha)u + b_{21}w, \tag{9}$$

which proves the following proposition.

PROPOSITION. *In a 2-dimensional real algebra with idempotent u, the coefficients b_{12} and b_{21} in Table 2 are invariant under change of basis from $\{u, v\}$ to $\{u, w\}$, as is $B = b_{12} + b_{21}$.* ■

The invariance of B and the fact that the complex numbers satisfy the condition $B \neq 0$ motivates the next definition.

DEFINITION. A 2-dimensional real division algebra given by Table 2 is *quasicomplex* if $B \neq 0$.

Real 2-dimensional quasicomplex division algebras (including $\mathbb{C}$) are classified in the next theorem.

THEOREM 8. *Let $\mathcal{A}$ and $\mathcal{A}'$ be quasicomplex 2-dimensional real division algebras, each with one idempotent, presented by multiplication tables T and T' in the form of Table 4 on the bases $\{u, v\}$ and $\{u', v'\}$, respectively. Let T'' be the table for $\mathcal{A}'$ based on $\{u', -v\}$. Then $\mathcal{A}$ and $\mathcal{A}'$ are isomorphic if and only if the coefficients of T are identical to those of T' or T''.*

Proof. If $\phi: \mathcal{A} \to \mathcal{A}'$ is an isomorphism, then $\phi(u) = u'$ since the algebras have only one idempotent, and $(\phi(v))^2 = \pm u'$. By the uniqueness of v' up to sign, $\phi(v) = \pm v'$. Since $\{u, v\}$ and $\{u', \phi(v)\}$ are bases, the tables must have the same coefficients. ■

Another approach to the simplification of Table 2 is to find a basis so that the table will resemble the complex numbers in having $a_{12} = a_{21} = 0$. This will be particularly useful if the algebra is not quasicomplex, but has independent interest. Such a basis change can be studied as before by investigating the effect on Table 2, but there is another way. A vector $w = \alpha u + \beta v$, $\beta \neq 0$ for which $u \cdot w = k_1 w$ and $w \cdot u = k_2 w$ is a solution to two simultaneous eigenvector problems. In fact, let R_u and L_u be the linear transformations on $\mathbb{R}^2$ defined by $R_u(z) = u \cdot z$ and $L_u(z) = z \cdot u$. Then w must satisfy

$$R_u(w) = k_1 w,$$
$$L_u(w) = k_2 w.$$

The matrices for R_u and L_u with respect to the basis $\{u, v\}$ are respectively,

$$\begin{bmatrix} 1 & a_{12} \\ 0 & b_{12} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & a_{21} \\ 0 & b_{21} \end{bmatrix}$$

with eigenvalues 1, corresponding to the eigenvector u, $b_{12} = k_1$, and $b_{21} = k_2$ corresponding to w. The simultaneous eigenvector equations are

$$\begin{cases} (1 - b_{12})\alpha + a_{12}\beta = 0 \\ (1 - b_{21})\alpha + a_{21}\beta = 0 \end{cases}, \tag{10}$$

which will have nontrivial solutions if and only if the determinant

$$\begin{vmatrix} 1-b_{12} & a_{12} \\ 1-b_{21} & a_{21} \end{vmatrix} = A_2 - (a_{12} - a_{21})$$

is zero. The solvability of the simultaneous eigenvector problem is independent of the basis vector v in Table 2, so if $A_2 = a_{12} - a_{21}$ for one basis as in Table 2, the same relation holds for any other basis as in Table 2. This invariance makes possible the following definition, suggested by the fact that for a commutative algebra $A_2 = 0 = a_{12} - a_{21}$.

DEFINITION. A 2-dimensional real division algebra given by Table 2 is *quasicommutative* if $A_2 = a_{12} - a_{21}$, where

$$A_2 = \begin{vmatrix} a_{12} & b_{12} \\ a_{21} & b_{21} \end{vmatrix}.$$

THEOREM 9. *A quasicommutative 2-dimensional real division algebra given by Table 2 in which b_{12} and b_{21} are not both 1 has a basis $\{u, v\}$ with multiplication table*

TABLE 5

$\cdot$	u	v
u	u	$b_{12}v$
v	$b_{21}v$	$a_{22}u + b_{22}v$

where $a_{22} = \pm 1$.

Proof. As we have seen, quasicommutativity guarantees the existence of the simultaneous eigenvector, and since either $b_{12} \neq 1$ or $b_{21} \neq 1$, there is a solution of (10) with $\beta \neq 0$. (If b_{12} or b_{21} is 1, it follows immediately from the definition of quasicommutativity that the corresponding a_{12} or a_{21} is zero.) Since $a_{22} \neq 0$ and v is an eigenvector, the pair $\{u, v/\sqrt{|a_{22}|}\}$ is a basis of eigenvectors whose multiplication table has $a_{22} = \pm 1$. ∎

Here is an example.

EXAMPLE 6.

$\cdot$	u	v
u	u	v
v	$-v$	$u+v$

The classification of nonquasicomplex algebras begins with the quasicommutative case.

THEOREM 10. *Let $\mathcal{Q}$ and $\mathcal{Q}'$ be nonquasicomplex, quasicommutative 2-dimensional real division algebras. Then each has a unique multiplication table in the form of Table 5. The algebras $\mathcal{Q}$ and $\mathcal{Q}'$ are isomorphic if and only if the coefficients in these tables are identical.*

Proof. Since $B = b_{12} + b_{21} = 0$, b_{12} and b_{21} cannot both equal 1. Thus, by Theorem 9, $\mathcal{Q}$ and $\mathcal{Q}'$ have Table 5 presentations. These are unique because the idempotent and the normalized solution to the eigenvector problem are unique. Since an isomorphism must preserve both, the coefficients agree. ∎

Note that a nonquasicomplex, quasicommutative division algebra is not commutative, since $b_{12} = b_{21} \neq 0$ is ruled out.

The only remaining algebras are those which are neither quasicommutative nor quasicomplex. Apparently, the best Table 2 simplification available is to make $a_{12} = a_{21} = 1$.

THEOREM 11. *A nonquasicommutative, nonquasicomplex 2-dimensional real division algebra with exactly one idempotent has a unique multiplication table in the form of Table 6.*

TABLE 6

$\cdot$	u	v
u	u	$u + b_{12}v$
v	$u - b_{12}$	$a_{22}u + b_{22}v$

Two such algebras are isomorphic if and only if the respective coefficients in Table 6 for each algebra are identical.

Proof. Setting the coefficients of u in equations (8) and (9) equal to 1 and noting that $b_{21} = -b_{12}$, we obtain

$$\begin{cases} (1 - b_{12})\alpha + a_{12}\beta = 1 \\ (1 + b_{12})\alpha + a_{21}\beta = 1 \end{cases}. \tag{11}$$

Since the algebra is not quasicommutative,

$$\begin{vmatrix} 1 - b_{12} & a_{12} \\ 1 + b_{12} & a_{21} \end{vmatrix} \neq 0$$

so (11) has a unique solution. By Cramer's Rule, $\beta = 0$ in this solution if and only if $b_{12} = 0$. But if $b_{12} = 0$, then $u \cdot (a_{12}u - v) = 0$, so that $a_{12}u - v = 0$ by (iii'), contradicting the fact that $\{u, v\}$ is a basis. Thus, this unique solution will yield Table 6. Since the table is unique, the isomorphism result is immediate. ■

Here is an example.

EXAMPLE 7.

$\cdot$	u	v
u	u	$u + v$
v	$u - v$	$2u + v$

5. Summary. Given a 2-dimensional real division algebra, one may select a basis so that the multiplication table assumes one of the following four forms:

TABLE 3
(Two or three idempotents)

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	v

TABLE 4
(Quasicomplex; includes $\mathbb{C}$)

$\cdot$	u	v
u	u	$a_{12}u + b_{12}v$
v	$a_{21}u + b_{21}v$	$\pm u$

TABLE 5
(Quasicommutative, not quasicomplex)

$\cdot$	u	v
u	u	$b_{12}v$
v	$b_{21}v$	$\pm u + b_{22}v$

TABLE 6
(Neither quasicommutative nor quasicomplex)

$\cdot$	u	v
u	u	$u + b_{12}v$
v	$u - b_{12}v$	$a_{22}u + b_{22}v$

Tables 5 and 6 are uniquely determined by the algebra, and Table 4 has only a sign ambiguity in v. A two-idempotent algebra is uniquely determined by Table 3 with $a_{12} + a_{21} = 1$. A three-idempotent algebra has 3 tables like Table 3, but each is readily obtained from any other. Note that these tables do *not* automatically define division algebras by virtue of their form. In fact, setting $b_{12} = 0$ in any one of them results in an algebra with zero divisors. It is therefore necessary to verify that condition (3) is satisfied.

The classification assigns algorithmically to every 2-dimensional real division algebra a canonical multiplication table by successively examining three properties: number of idempotents, quasicomplexity, and quasicommutativity. It is noteworthy that for some division algebras there exist tables of more than one canonical form. For example, a nonquasicommutative algebra may be presented by Table 4 as well as Table 6, provided $b_{12} \neq b_{21}$.

6. Concluding Remarks. Although all our results depend upon first locating an idempotent, it is easy to generalize Theorem 3 so that it applies directly to Table 1.

THEOREM 12. *An algebra described by Table* 1 *is a division algebra if and only if*

$$(A_4 - A_1)^2 < 4A_2A_3,$$

where A_1 and A_2 are as before and

$$A_3 = \begin{vmatrix} a_{11} & b_{11} \\ a_{12} & b_{12} \end{vmatrix} \quad A_4 = \begin{vmatrix} a_{11} & b_{11} \\ a_{22} & b_{22} \end{vmatrix}. \blacksquare$$

The tables for a three-idempotent algebra can also be determined without actually finding any idempotents. A slightly deeper investigation of the rotation described in Theorem 2 yields functions

$$F(x) = a_{12}x^3 + (a_{11} - a_{22} - b_{12})x^2 - (a_{21} + b_{11} - b_{22})x + b_{21},$$
$$G(x) = a_{21}x^3 + (a_{11} - a_{22} - b_{21})x^2 - (a_{12} + b_{11} - b_{22})x + b_{12},$$

which can be used to generate table coefficients as follows. Let r_i, $i = 1,2,3$ denote the three real roots of $g(1, y) = 0$. Let

$$r_{1i} = \frac{F(r_i)}{f(1, r_i)}, \quad r_{2i} = \frac{G(r_i)}{f(1, r_i)}.$$

Then all tables based on idempotents are given by

$\cdot$	u	v
u	u	$r_{1i}u + r_{2j}v$
v	$r_{2i}u + r_{1j}v$	v

It is also possible to give coordinate-free characterizations of two important concepts in this paper.

PROPOSITION. *An algebra $\mathcal{A}$ is* not *quasicomplex if and only if there exists a real-valued function $k: \mathcal{A} \to \mathbb{R}$ such that*

$$uv + vu = k(v)u$$

for all $v \in \mathcal{A}$. $\blacksquare$

PROPOSITION. *An algebra $\mathcal{A}$ with idempotent u is quasicommutative if and only if*

$$(uv)u = u(vu)$$

for all $v \in \mathcal{A}$. $\blacksquare$

Finally, we note that the discussion preceding Theorem 3 implies that $\mathbb{C}^2$ is never a division

algebra since by the Fundamental Theorem of Algebra, the quadratic form Q will always have nontrivial zeros. In particular, no multiplication on $\mathbb{C}^2$ will make it the quaternions.

7. Historical Note. The study of linear algebras was initiated by Benjamin Peirce, who by 1870 had worked out 115 multiplication tables [**7**]. In 1908 J. H. M. Wedderburn [**10**, p. 111] gave what we believe to be the first example of a 2-dimensional division algebra without identity (although it was over the integers modulo 2, not $\mathbb{R}$). In 1954 Segre [**8**] used Bezout's Theorem to prove Theorem 2 for all finite-dimensional real division algebras. In 1958 Luchian [**4**] classified 2-dimensional real algebras with divisors of zero. Thus, this paper together with ours gives a classification of all 2-dimensional real algebras. In 1960 Markus [**5**] classified commutative 2-dimensional real algebras to help in the study of quadratic differential equations. Finally, in 1970 Wallace [**9**] gave a classification of 2-dimensional power associative real linear algebras.

The research for this project was supported (in part) by a grant from the Faculty Development Fund of the University of Michigan-Flint.

References

1. S. C. Althoen and J. F. Weidner, Real division algebras and Dickson's construction, this MONTHLY, 85 (1978) 368–371.

2. M. Hall, Jr., The Theory of Groups, Chelsea, NY, 1976, p. 7.

3. I. N. Herstein, Topics in Algebra, 2nd ed., Wiley, NY, 1964, pp. 368–371.

4. T. Luchian, A classification of linear algebras of order 2, with divisors of zero, An. Stiint. Univ. "Al. I. Cuza" Iasi Sect. I, IV (1958), 21–37.

5. L. Markus, Quadratic Differential Equations and Nonassociative Algebras, Contributions to the Theory of Nonlinear Oscillations, Princeton, 1960, pp. 185–213.

6. R. S. Palais, The classification of real division algebras, this MONTHLY, 75 (1968) 366–368.

7. B. Peirce, Linear associative algebras, Amer. J. Math., 4 (1881) 97–215; addenda, 216–229.

8. B. Segre, La teoria delle algebre ed alcune questioni di realtà, Rend. Mat. e Appl. serie 5, 13 (1954–5), 157–188.

9. E. W. Wallace, Two-dimensional power-associative algebras, Math. Mag., 43 (1970) 158–162.

10. J. H. M. Wedderburn, On hypercomplex numbers, Proc. London Math. Soc., Series 2, 6 (1908) 77–118.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

A SIMPLE PROOF OF FERMAT'S TWO-SQUARE THEOREM

JOHN A. EWELL

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

1. Introduction. As observed by Hardy and Wright [**2**, p. 300], the problem of characterizing the natural numbers which are representable by sums of two squares reduces ultimately to proof of the following classical theorem of Fermat.

THEOREM 1. *Any rational prime p of the form $4m + 1$ can be expressed as a sum of two squares.*

The adjective "rational" is here required to distinguish between the domain of "ordinary" integers and the domain of Gaussian integers, a setting in which Theorem 1 can be proved. However, it is certainly desirable to construct proofs of this important theorem which are independent of the theory of Gaussian integers. Accordingly, we propose in this note to deduce

algebra since by the Fundamental Theorem of Algebra, the quadratic form Q will always have nontrivial zeros. In particular, no multiplication on $\mathbb{C}^2$ will make it the quaternions.

7. Historical Note. The study of linear algebras was initiated by Benjamin Peirce, who by 1870 had worked out 115 multiplication tables [7]. In 1908 J. H. M. Wedderburn [**10**, p. 111] gave what we believe to be the first example of a 2-dimensional division algebra without identity (although it was over the integers modulo 2, not $\mathbb{R}$). In 1954 Segre [**8**] used Bezout's Theorem to prove Theorem 2 for all finite-dimensional real division algebras. In 1958 Luchian [**4**] classified 2-dimensional real algebras with divisors of zero. Thus, this paper together with ours gives a classification of all 2-dimensional real algebras. In 1960 Markus [**5**] classified commutative 2-dimensional real algebras to help in the study of quadratic differential equations. Finally, in 1970 Wallace [**9**] gave a classification of 2-dimensional power associative real linear algebras.

The research for this project was supported (in part) by a grant from the Faculty Development Fund of the University of Michigan-Flint.

References

1. S. C. Althoen and J. F. Weidner, Real division algebras and Dickson's construction, this MONTHLY, 85 (1978) 368–371.

2. M. Hall, Jr., The Theory of Groups, Chelsea, NY, 1976, p. 7.

3. I. N. Herstein, Topics in Algebra, 2nd ed., Wiley, NY, 1964, pp. 368–371.

4. T. Luchian, A classification of linear algebras of order 2, with divisors of zero, An. Stiint. Univ. "Al. I. Cuza" Iasi Sect. I, IV (1958), 21–37.

5. L. Markus, Quadratic Differential Equations and Nonassociative Algebras, Contributions to the Theory of Nonlinear Oscillations, Princeton, 1960, pp. 185–213.

6. R. S. Palais, The classification of real division algebras, this MONTHLY, 75 (1968) 366–368.

7. B. Peirce, Linear associative algebras, Amer. J. Math., 4 (1881) 97–215; addenda, 216–229.

8. B. Segre, La teoria delle algebre ed alcune questioni di realtà, Rend. Mat. e Appl. serie 5, 13 (1954–5), 157–188.

9. E. W. Wallace, Two-dimensional power-associative algebras, Math. Mag., 43 (1970) 158–162.

10. J. H. M. Wedderburn, On hypercomplex numbers, Proc. London Math. Soc., Series 2, 6 (1908) 77–118.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

A SIMPLE PROOF OF FERMAT'S TWO-SQUARE THEOREM

JOHN A. EWELL

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

1. Introduction. As observed by Hardy and Wright [**2**, p. 300], the problem of characterizing the natural numbers which are representable by sums of two squares reduces ultimately to proof of the following classical theorem of Fermat.

THEOREM 1. *Any rational prime p of the form $4m + 1$ can be expressed as a sum of two squares.*

The adjective "rational" is here required to distinguish between the domain of "ordinary" integers and the domain of Gaussian integers, a setting in which Theorem 1 can be proved. However, it is certainly desirable to construct proofs of this important theorem which are independent of the theory of Gaussian integers. Accordingly, we propose in this note to deduce

Theorem 1 as a simple consequence of well-known partition identities. Details of the proof are supplied in Section 2. In this the leading role is played by the celebrated Gauss-Jacobi triple-product identity:

$$\prod_{n=1}^{\infty}(1-x^{2n})(1+ax^{2n-1})(1+a^{-1}x^{2n-1})=\sum_{n=-\infty}^{\infty}a^{n}x^{n^{2}}, \tag{1}$$

valid for each pair of complex numbers a, x such that $a \neq 0$ and $|x| < 1$. For an accessible proof of (1) see this MONTHLY [**1**, pp. 270–272]. As a matter of fact, we do not require the full force of (1). Rather, we need the following special case which results from the substitution $a \to -1$: viz.,

$$\prod_{1}^{\infty}(1-x^{n})(1-x^{2n-1})=1+2\sum_{1}^{\infty}(-1)^{n}x^{n^{2}}. \tag{2}$$

2. Proof of Theorem 1. For the sake of brevity let $F(x)$ denote the right side of (2), then take the logarithmic derivative and multiply the resulting identity by x to get:

$$\begin{aligned}
x\frac{F'(x)}{F(x)} &= xD_x\{\log F(x)\}\\
&= xD_x\left\{\sum_{1}^{\infty}\log(1-x^{n})+\sum_{1}^{\infty}\log(1-x^{2n-1})\right\}\\
&= -\sum_{1}^{\infty}\frac{nx^{n}}{1-x^{n}}-\sum_{1}^{\infty}\frac{(2n-1)x^{2n-1}}{1-x^{2n-1}}\\
&= -\sum_{1}^{\infty}x^{n}\sum_{d\mid n}d-\sum_{1}^{\infty}x^{n}\sum_{\substack{d\mid n\\ d\text{ odd}}}d.
\end{aligned}$$

The following definition naturally simplifies our discussion.

DEFINITION. For each positive integer n, $\sigma(n)$ denotes the sum of all positive divisors of n, the nonnegative integer $b(n)$ denotes the exponent of the highest power of 2 dividing n, and $0(n)$ is then defined by the equation $n = 2^{b(n)}0(n)$. Hence, $0(n)$ is odd. For convenience we define an auxiliary arithmetical function w by the equation $w(n) = \sigma(n) + \sigma(0(n))$.

We now state our result, as follows:

$$\left\{\sum_{1}^{\infty}w(n)x^{n}\right\}F(x)=-xF'(x),$$

or equivalently,

$$\left\{\sum_{1}^{\infty}w(n)x^{n}\right\}\left\{1+2\sum_{1}^{\infty}(-1)^{n}x^{n^{2}}\right\}=-2\sum_{1}^{\infty}(-1)^{n}n^{2}x^{n^{2}}.$$

Expanding the left side of the foregoing identity we have:

$$\sum_{1}^{\infty}x^{n}\left\{w(n)+2\sum_{j=1}(-1)^{j}w(n-j^{2})\right\}=-2\sum_{1}^{\infty}(-1)^{n}n^{2}x^{n^{2}}.$$

On the left side of this identity summation of the inner sum is extended over all values of j for which $n - j^2$ is positive. Equating coefficients of like powers we then have:

$$w(n)+2\sum_{j=1}(-1)^{j}w(n-j^{2})=\begin{cases}2(-1)^{r+1}r^{2}, & \text{if } n=r^{2},\\ 0, & \text{otherwise.}\end{cases}$$

For n odd, $w(n) = \sigma(n) + \sigma(0(n)) = 2\sigma(n)$. Hence, for such n we divide the foregoing

recurrence by 2 to get:

(3)

$$\sigma(2m+1) - \sum_{j=1} w\left(2m+1-(2j-1)^2\right) + 2\sum_{j=1} \sigma\left(2m+1-(2j)^2\right) = \begin{cases} r^2, & \text{if } 2m+1=r^2, \\ 0, & \text{otherwise.} \end{cases}$$

Now let there be given a prime $p = 4m+1$, for some integer m. Hence $\sigma(p) = p+1 = 4m + 2$, and recurrence (3) becomes:

$$4m+2-\sum_1 w\left(4m+1-(2j-1)^2\right) + 2\sum_1 \sigma\left(4m+1-(2j)^2\right) = 0,$$

or equivalently,

$$2m+1-\sum_1 \frac{w\left(4m+1-(2j-1)^2\right)}{2} + \sum_1 \sigma\left(4m+1-(2j)^2\right) = 0.$$

Next

$$w(n) = \sigma\left(2^{b(n)}0(n)\right) + \sigma(0(n)) = 2^{b(n)+1}\sigma(0(n)),$$

owing to the multiplicativity of σ. Hence, 2 divides $w(n)$; and, for n even, 4 divides $w(n)$. From this it follows that the sum $\Sigma\sigma(4m+1-(2j)^2)$ is odd, whence at least one of the summands, say $\sigma(4m+1-(2k)^2)$, is odd. But, as is well known (and easy to prove), $\sigma(n)$ is odd if and only if n is a square or twice a square. In the present case we must have: $4m+1-(2k)^2 = (2i+1)^2$, for some integer i. Thus,

$$p = (2k)^2 + (2i+1)^2.$$

REMARKS. In [**3**, pp. 446–448] Uspensky and Heaslet discuss an arithmetical function T which is related to our function w by: $2T(n) = w(n)$. These authors, following Liouville, then establish the desired recursive formula for T; and, as an application of the recurrence they prove Fermat's theorem by use of an argument quite similar to ours. However, their discussion depends on a very lengthy prior discussion of the methods of Liouville, whereas our discussion begins with the triple-product identity and is thereafter self-contained.

References

1. J. A. Ewell, An easy proof of the triple-product identity, this MONTHLY, 88 (1981) 270–272.

2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960.

3. J. V. Uspensky and M. A. Heaslet, Elementary Number Theory, 1st ed., McGraw-Hill, New York, 1939.

NECESSARY AND SUFFICIENT CONDITIONS FOR OSCILLATIONS

G. LADAS, Y. G. SFICAS,* AND I. P. STAVROULAKIS**

Department of Mathematics, University of Rhode Island, Kingston, R.I. 02881

1. Introduction. Our aim in this paper is to obtain a necessary and sufficient condition under which all solutions of the retarded differential equation

$$x'(t) + \sum_{i=1}^{n} p_i x(t-\tau_i) = 0, \quad t \geqslant 0 \tag{1}$$

oscillate, where p_i are positive numbers and τ_i nonnegative numbers, $i = 1, 2, \ldots, n$.

*On sabbatical leave from the University of Ioannina, Greece.

**On leave from the University of Ioannina, Greece.

Retarded differential equations provide mathematical models for physical systems in which the rate of change of the system depends not only on its present stage, but also on its past history. Such equations appear in biology, control theory, ecology, economics, number theory, physics, etc.

The Oscillation Theory of retarded differential equations has been extensively developed during the past few years. See, for example, **[1]**–**[8]** and the references cited therein. Of particular importance has been the study of oscillations which are caused by the retarded arguments. Note that when all delays τ_i in Equation (1) are zero, that is, in the case of a first order ordinary differential equation, the nontrivial solutions are not oscillatory. However, for retarded differential equations, under appropriate hypotheses, it is possible to have nontrivial oscillatory solutions. The behavior of solutions is radically altered by the introduction of delays.

It is known that in the case of stability the roots of the characteristic equation determine the stability character of the solutions. Our conjecture was that the oscillatory character of the solutions was also determined by the roots of the characteristic equation. This conjecture is in fact true and the proof requires only elementary tools.

Our main result is the following theorem.

THEOREM. *Consider the retarded differential equation*

$$x'(t) + \sum_{i=1}^{n} p_i x(t-\tau_i) = 0, \quad t \geqslant 0 \tag{1}$$

where p_i are positive numbers and τ_i nonnegative numbers, $i = 1,\ldots,n$. Then all solutions of (1) *oscillate if and only if*

$$-\lambda + \sum_{i=1}^{n} p_i e^{\lambda\tau_i} > 0, \quad \text{for all } \lambda > 0. \tag{2}$$

As is customary, a function is said to *oscillate* if it has arbitrarily large zeros. We shall assume without loss of generality that $\tau_n = \max\{\tau_1,\ldots,\tau_n\}$; *solutions* of (1) are then continuous functions x defined on $[-\tau_n, \infty)$ that satisfy (1). As usual, we shall use the term "*eventually*" to mean "for some $T \in (0,\infty)$ and all $t \in [T,\infty)$".

2. Proof of Main Result. To prove the theorem we need the following lemma.

LEMMA. *If the continuous positive function x defined on $[-\tau,\infty)$ is a solution of the retarded differential inequality*

$$x'(t) + px(t-\tau) \leqslant 0, \quad t \geqslant 0 \tag{3}$$

where p and τ are positive numbers, then

$$(p\tau/2)^2 x(t-\tau) \leqslant x(t), \quad t \geqslant 3\tau/2.$$

Proof. For given $s \geqslant \tau$, we integrate both sides of (3) from s to $s+\tau/2$; using the fact that x is decreasing on $[0,\infty)$, we find that

$$-x(s) + (p\tau/2)x(s-\tau/2) \leqslant x(s+\tau/2) - x(s) + (p\tau/2)x(s-\tau/2) \leqslant 0, \quad s \geqslant \tau. \tag{4}$$

For given $t \geqslant 3\tau/2$ we apply (4) to $s = t - \tau/2$ and to $s = t$, and find

$$x(t-\tau/2) \geqslant (p\tau/2)x(t-\tau) \quad \text{and} \quad x(t) \geqslant (p\tau/2)x(t-\tau/2), \quad t \geqslant 3\tau/2;$$

the assertion follows by combining these inequalities.

Proof of the Theorem. Assume first that (2) does not hold. We may then choose $\lambda_0 > 0$ such that

$$-\lambda_0 + \sum_{i=1}^{n} p_i e^{\lambda_0\tau_i} = 0.$$

But then x, defined on $[-\tau_n,\infty)$ by $x(t) = e^{-\lambda_0 t}$, is a nonoscillatory solution of (1).

Assume conversely, that not all solutions of (1) oscillate. There exists, therefore, a solution that is eventually positive; by (1), such a solution is eventually decreasing. Since every left-translate of a solution of (1) is also a solution, we may choose a decreasing positive solution x of (1), and this shall remain fixed throughout the proof.

When all τ_i are 0, (2) obviously does not hold. We shall therefore assume that $\tau_n = \max\{\tau_1, \dots, \tau_n\} > 0$. We define the set

$$\Lambda = \{\lambda > 0 \,|\, x'(t) + \lambda x(t) < 0 \text{ eventually}\}.$$

From (1) we have

$$x'(t) + p_n x(t - \tau_n) \leqslant 0, \quad t \geqslant 0. \tag{5}$$

On the one hand, since x is decreasing, if follows from (5) that

$$x'(t) + p_n x(t) < 0, \quad t \geqslant 0,$$

so that $p_n \in \Lambda$. On the other hand, it follows from (5) by the Lemma that

$$x(t - \tau_n) \leqslant \left(\frac{2}{p_n \tau_n}\right)^2 x(t), \quad t \geqslant 3\tau_n/2.$$

Since x is decreasing, this implies

$$\begin{aligned} 0 = x'(t) + \sum_{i=1}^{n} p_i x(t - \tau_i) &\leqslant x'(t) + \sum_{i=1}^{n} p_i x(t - \tau_n) \\ &\leqslant x'(t) + \left(\frac{2}{p_n \tau_n}\right)^2 \sum_{i=1}^{n} p_i x(t), \quad t \geqslant 3\tau_n/2. \end{aligned}$$

Therefore $(2/p_n\tau_n)^2 \sum_{i=1}^{n} p_i$ is an upper bound of Λ. Since Λ is nonempty and bounded, we may set $\lambda_0 = \sup \Lambda$.

Let $\lambda \in \Lambda$ be given, and define y on $[-\tau_n, \infty)$ by $y(t) = x(t)e^{\lambda t}$. Then there is a suitable $T \in (0, \infty)$ such that

$$y'(t) = (x'(t) + \lambda x(t))e^{\lambda t} < 0, \quad t \geqslant T$$

and therefore y is decreasing on $[T, \infty)$. It follows that

$$\begin{aligned} 0 = x'(t) + \sum_{i=1}^{n} p_i x(t - \tau_i) &= x'(t) + \sum_{i=1}^{n} p_i y(t - \tau_i) e^{-\lambda(t - \tau_i)} \\ > x'(t) + \sum_{i=1}^{n} p_i y(t) e^{-\lambda(t - \tau_i)} &= x'(t) + \sum_{i=1}^{n} p_i e^{\lambda \tau_i} x(t), \quad t \geqslant T + \tau_n. \end{aligned}$$

This shows that $\sum_{i=1}^{n} p_i e^{\lambda \tau_i} \in \Lambda$ and therefore $\sum_{i=1}^{n} p_i e^{\lambda \tau_i} \leqslant \lambda_0$. Since $\lambda \in \Lambda$ was arbitrary, we conclude that

$$\sum_{i=1}^{n} p_i e^{\lambda_0 \tau_i} \leqslant \lambda_0$$

and therefore (2) does not hold.

The proof of the theorem is complete.

Note. After we had derived the above result (see [4]), we discovered that the same theorem was also proved by Tramov [**8**]. An equivalent version of the theorem has also been proved by Hunt and Yorke [**1**]. The three proofs are different.

References

1. B. R. Hunt and J. A. Yorke, When all solutions of $x' = -\sum_{i=1}^{n} q_i(t)x(t - T_i(t))$ oscillate (preprint).

2. T. Kusano, On even order functional differential equations with advanced and retarded arguments, J.

Differential Equations, 45 (1982) 75–84.

3. G. Ladas, Sharp conditions for oscillations caused by delays, Applicable Anal., 9 (1979) 93–98.

4. G. Ladas, Y. G. Sficas and I. P. Stavroulakis, Necessary and sufficient conditions for oscillations, Abstracts Amer. Math. Soc., 3, 7 (1982) 597.

5. G. Ladas and I. P. Stavroulakis, Oscillations caused by several retarded and advanced arguments, J. Differential Equations, 44 (1982) 134–152.

6. V. N. Ševelo and N. V. Vareh, Asymptotic methods in the theory of nonlinear oscillations, Kiev "Naukova Dumka," 1979 (Russian).

7. Y. G. Sficas and V. A. Staikos, Oscillations of differential equations with deviating arguments, Funkcial. Ekvac., 19 (1976) 35–43.

8. M. I. Tramov, Conditions for oscillatory solutions of first order differential equations with a delayed argument, Izv. Vysš. Učebn. Zaved., Matematika 19, 3 (1975) 92–96.

R^3 IS THE UNION OF DISJOINT CIRCLES

ANDRZEJ SZULKIN
Department of Mathematics, University of Stockholm, 11385 Stockholm, Sweden

In a problem seminar held at the Royal Institute of Technology in Stockholm, Professor H. S. Shapiro asked for a solution of the problem of covering the three-dimensional Euclidean space R^3 by disjoint Jordan curves. (It can be shown that the two-dimensional Euclidean space cannot be covered in such a way.) In this note we shall construct a family of disjoint *circles* whose union is R^3. By a circle we mean a set isometric to $\{(x, y) \in R^2 : x^2 + y^2 = \rho^2\}$, where ρ is a positive real number.

Given $r \geqslant 0$, let $S_r = \{(x, y, z) \in R^3 : x^2 + y^2 + z^2 = r^2\}$. Then

$$R^3 = \bigcup_{r \geqslant 0} S_r.$$

Denote by C the union of the circles

$$\{(x, y, z) \in R^3 : (x - 4k - 1)^2 + y^2 = 1, z = 0\}, \quad k = 0, \pm 1, \pm 2, \dots .$$

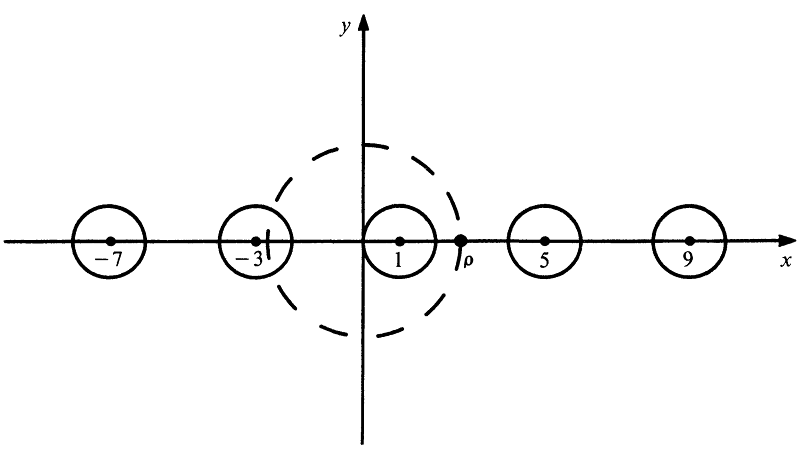


FIG. 1

One verifies readily that $C \cap S_r$ consists of two points for each $r > 0$ (see Fig. 1). Hence

$$R^3 - C = \bigcup_{r > 0} T_r,$$

where

$$T_r = S_r - (C \cap S_r) = S_r - \{\text{two points}\}.$$

Now one can remove a great circle C_r from each T_r, $r > 0$, in such a way that $T_r - C_r = T_r' \cup T_r''$, where T_r' and T_r'' are open hemispheres with one point missing. Thus it remains to cover all T_r' and T_r'' by disjoint circles. The covering can be obtained, e.g., by intersecting each T_r' and T_r'' with a family of planes as indicated by Fig. 2.

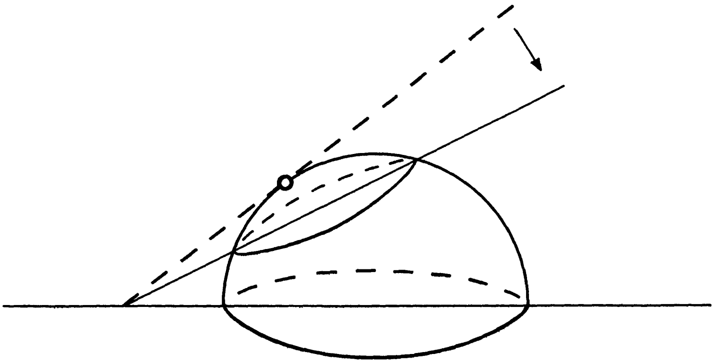


FIG. 2

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

A WAY OF TEACHING ABSTRACT ALGEBRA

HAYA FREEDMAN

Department of Mathematics, LSE, Houghton Street, London, England WC2A 2AE

> "We are informed about everything. We know nothing."
> (S. Bellow, *To Jerusalem and Back*.)

The lecture method is the most commonly used in teaching mathematics, in spite of its many shortcomings which are widely recognized by mathematicians and educators.

"A good lecture is usually systematic, complete, precise—and dull; it is a bad teaching instrument," writes P. R. Halmos [**1**]. "It is simplistic to suppose that people remember what they are told and understand the things that are explained to them clearly," states E. E. Moise [**2**].

This kind of criticism created the impetus to find a new way of teaching mathematics. During the last seven years, a method of teaching algebra, which does not depend solely on books and lectures, has been developed at this college. The essence of this method is to encourage students to play a more active part in their own education, in particular, to let them share responsibility with their teacher for selecting, preparing, and presenting the material taught in algebra. The purpose of this paper is to describe how this method works.

The students are introduced to this way of teaching gradually, in three stages. The first year students have, in addition to a traditional lecture in algebra, an informal seminar, which is done on a voluntary basis—no marks are awarded for it. Each participant has to read an extract from a paper or a book, or solve an interesting problem and lecture on it to the class. The listeners are encouraged to ask as many questions as they wish.

The second year seminar is a much more formal affair, and a whole term is devoted to it. The amount of work required to complete a second year project is considerably greater than that spent on a first year seminar—in addition to a couple of lectures, each student has to submit a written report, and this time, unlike the first year, marks are awarded for the project. Material for the projects is distributed at the end of the first year, so that the students can do some preliminary work during their summer vacation, and then complete their projects during term time. Each student is given one or two original papers and a list of prerequisites with references to accessible literature. After a short introductory talk, the student is packed off to do the work on his own. He or she is very welcome to return for help with any problems that may arise, but no advice is ever offered unless explicitly requested by the student. Each speaker is allowed three or four hours to lecture on his or her project. The student has to decide how to arrange the material: which of the proofs to give in full, and which to outline, or omit altogether. The student is also required to provide suitable examples and counterexamples. In addition, any material new to the class, or even only to some of its members, which is needed as a prerequisite for understanding the lecture, has to be explained adequately by the speaker. Following the lecture, each student has a tête-á-tête discussion with the instructor, in which both the lecture and the written report are scrutinized.

The third year courses are designed with the following objectives in mind: First the majority of lectures are to be prepared and delivered by students. Second, the whole course is constructed around a few focal points of interest. Third, students are instructed to use suitable original papers rather than books and lecture notes.

To make these points clearer a course on abelian groups, given in 1979 at this college, is described in some detail. Four students took this course. Each student was given a chapter from a book specially compiled for this course. Such a chapter included up to four original papers dealing with a specific topic in abelian groups, explanatory notes to cover difficult points, and exercises designed to fill in gaps in proofs, or to widen the scope of the papers. In addition, it contained an introduction and a summary with references for further study.

The first student was given the task of presenting criteria under which an abelian group is a direct sum of cyclic groups. His chapter included the following papers:

1. A. Kertesz, On the decomposibility of abelian p-groups into the direct sum of cyclic groups, Acta Math. Acad. Sci. Hungar., 3 (1952) 121–125.
2. R. Rado, A proof of the basis theorem for finitely generated abelian groups, J. London Math. Soc., 26 (1951) 74–75.
3. T. Szele, On a theorem of Pontrajagin, Acta Math. Acad. Sci. Hungar., 2 (1951) 121–125.
4. T. Szele, On direct sums of cyclic groups, Publ. Math. Debrecen, 2 (1952) 76–78.

On his own initiative he added to his reading list extracts from papers by R. Baer, H. Prüfer and L. Kulikov.

The second student lectured on basic subgroups. His chapter included papers by L. Fuchs, P. Hill, M. L. Kaloujnine and T. Szele. The other two talks were on pure subgroups and Ulm's theorem. The latter of these proved Ulm's theorem for countable abelian groups and for totally projective groups.

The four students worked as a group and had a great deal of discussion among themselves. Between them they gave about fifteen two-hour lectures and covered in their written projects about a dozen original papers. At the end of the academic year each student had to pass a written exam based on all of the four projects, excluding some of the more advanced material. For

example, the syllabus for the written exam included Ulm's theorem for countable abelian groups but excluded Walker's paper on Ulm's theorem for totally projective groups.

Each project, whether intended for a beginner or a more advanced student, should be selected with the utmost care. The following guide lines are used.

(a) The project must be related to what the student already knows, should know or desires to know.
(b) The educational needs of each student must be taken into consideration. In particular, weaker students should be offered easy projects, whereas gifted students should be stimulated to develop their intellectual powers and aesthetic perceptions to their full capacity.
(c) One must work with each student individually, and endeavour to answer cheerfully all the questions asked, whether it involves a lengthy explanation of a trivial matter, or requires the instructor to learn material new to him or her.
(d) Students must be given freedom to develop their own way of collecting, analyzing and evaluating any information they might need for their work.
(e) Since it is not always easy to forecast how well a student will take to the prescribed material, each project has to be constructed in two or three sections. The first section presents a mini project which grows to a more advanced project with the addition of further sections. In this way it is possible for the student to gain a sense of achievment and satisfaction even if only a part of the project is accomplished.

Our experience has shown that most students love this method of teaching and that they derive a great sense of satisfaction from their work. They tend to achieve better exam results than students who are taught in the traditional way. Furthermore, a higher percentage of graduates among those who have been exposed to seminars in algebra become interested in taking a higher degree in mathematics (usually specializing in algebra). Finally, since students are enabled to study at different levels, according to their own talents and interests, they become more forthcoming, and it is easier for the instructor to assess the potential of each student and identify his or her weak points.

The main disadvantage of this method is that it works effectively for small size classes only, say about a dozen students, although it may possibly be adaptable to larger classes. It probably means additional work for the teacher, although not as much as might seem, since one obtains a great deal of help from the students. For example, one of the first year students discovered, after giving his seminar, that he loved teaching, and he organized a special revision class to help some of his classmates. On another occasion, a Master of Philosophy student volunteered to supervise the project of a second year student. She was convinced that this was a good way of learning new material in which she happened to be interested at that time. At the end of the academic year students are asked for their comments on the teaching method. Here are a few extracts from their essays.

"First and foremost, the seminar is taught by students; in short, it is a place where discussions and group participation are encouraged. Second, the seminar exposes the student to a dozen topics that could not be possibly be covered in lectures. Whereas the lecture provides one with a sound algebraic base, the seminar discovers areas in mathematics that are beyond the definitions and basic theorems. Third, the situation of students as professor is fruitful not only because the transformation of learner into teaching is rewarding (yet demanding) but also because, as teacher, the student will learn his or her subject matter like no other way." (S. Fine, a general course student reading a first year algebra course.)

"I feel that every student should be made to read a number of research papers. This exercise gives a new perspective on how human knowledge grows, something a text book can rarely do. Finally, the most beneficial aspect of the seminar system is that it improves a person's ability to communicate information. In my own case, I learned a great deal on how to present material—what

items to stress, what details to omit, etc. This was a most rewarding experience, especially since it increased my confidence in my own abilities." (M. Khan, reading for a Master of Science Degree.)

"One of the most important aspects for me is the sense of satisfaction experienced on completion of the seminar." (P. A. Smith, third year.)

Finally, two comments are in order. First, no special magic attaches to abstract algebra. It was chosen for this experiment purely for personal reasons, since it happened to be the author's field of interest. The method could equally be applied to the teaching of other subjects in mathematics.

Second, it is not necessary that a subject be taught for three years. Experience has shown that students who took only first and second year algebra achieved many of the same benefits as those who did abstract algebra for three years.

References

1. The problem of learning to teach, this MONTHLY, 82 (1975) 466–476.
2. Ibid.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, typed and in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by March 31, 1984. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowlegdement).

E 3019. *Proposed by Chen-Te Yen, Chung-Yuan Christian University, Taiwan.*

Find all solutions to the equations (1) $1 + 3^a = 7^b + 3^c$ and (2) $1 + 5^a = 7^b + 5^c$ in integers a, b, c.

E 3020. *Proposed by Clark Kimberling, University of Evansville.*

Suppose ABC is a nonisosceles triangle. Find three hyperbolas concurrent in a point P such that triangles APB, APC, and BPC all have the same perimeter. (*) How does this common perimeter compare with that of ABC?

E 3021. *Proposed by Geng-Zhe Chang and Edward T. H. Wang, Wilfrid Laurier University.*

Let

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k}^2 (1+x)^k (1-x)^{n-k}.$$

Express $p_n(x)$ as an explicit function of $1 - x^2$.

E 3022. *Proposed by John Sadowsky, Columbia, MD.*

Show that, for any $\alpha > 0$ and any positive integer N,

$$\sum_{k=1}^{N} (-1)^{k-1} \binom{N}{k} \frac{k}{1+(k-1)\alpha} = \prod_{k=1}^{N-1} \left(\frac{k+1}{k+\alpha^{-1}} \right).$$

E 3023. *Proposed by J. G. Mauldon, Amherst College.*

You are given an accurate indicating spring scale (not a balance) and seven coins, each of which weighs either x or y, where x and y are unknown. In five weighings, determine the weight of each coin.

E 3024. *Proposed by Eugene M. Luks, Bucknell University.*

Let a, b be integers with $b > 0$. Prove that there are infinitely many positive integers n with the property that if p is a prime divisor of $n^b + a$, then p is also a divisor of $k^b + a$ for some integer k with $|k|^b < n$. (The case $b = 2$, $a = 3$ was problem B–3 on the 1981 William Lowell Putnam Competition.)

SOLUTIONS OF ELEMENTARY PROBLEMS

Primes, Ratios, and Bernoulli Numbers

E 2890 [1981, 444]. *Proposed by Barry J. Powell, Kirkland, Washington.*

Let $p > 3$ be a prime for which $2p - 1 = q$ is also prime. Let $B_{2m} = N_{2m}/D_{2m}$ represent the $2m$th Bernoulli number, where N_{2m} is the numerator and D_{2m} is the denominator (in usual notation). Prove that $|N_{2p-2}| > |N_{2p}|$.

*This suggests that there are infinitely many positive integers m for which $|N_{2m-2}| > |N_{2m}|$. Can this be proved?

Solution by J. S. Frame, Michigan State University. The primes p and $q = 2p - 1$ have the form $p = 6n + 1$, $q = 12n + 1$, $n > 0$. For $n = 1$,

$$\left| \frac{N_{12}}{N_{14}} \right| = \frac{691}{7} > 10\pi^2.$$

For $n > 1$, we show that $|N_{2p-2}| > 8981|N_{2p}|$, and conjecture that the ratio $|(N_{2p-2})/N_{2p}|$ is unbounded. By the von Staudt-Clausen theorem, the denominator D_{2m} is the product of the distinct primes r such that $(r-1)|2m$. Since $p + 1$ and $2p + 1$ are divisible respectively by 2 and 3, $D_{2p} = 6$. Also

$$D_{2p-2} = D_{12n} = 2 \cdot 3 \cdot 5 \cdot R \cdot p \cdot q = 5RpqD_{2p}$$

where $5R$ is the product of all primes r between 4 and $6n$ such that $(r-1)|12n$. Next, from the zeta function relation

$$\zeta(2m) = \sum_{n=1}^{+\infty} n^{-2m} = |B_{2m}|(2\pi)^{2m}/2(2m)! > \zeta(2m+2)$$

we conclude that $|B_{2p-2}/B_{2p}| > 2\pi^2/pq$, and hence $|N_{2p-2}/N_{2p}| > 10\pi^2 R$. Since 25 is not prime, $n \neq 2$. For $n > 2$, we note that $(7 \cdot 13)|R$, so

$$|N_{2p-2}/N_{2p}| > 910\pi^2 > 8981$$

for $p > 7$, and $2p - 1$ prime. The values of $R/91$ for the first ten prime pairs p, q corresponding to $n = 1, 3, 5, 6, 13, 16, 23, 26, 33, 35$ are $1/91, 1, 11, 19, 53, 17, 47, 53 \cdot 79, 19 \cdot 23 \cdot 37 \cdot 67, 11 \cdot 29 \cdot 31 \cdot 43 \cdot 61 \cdot 71$. It appears probable that R and the ratio $|N_{2p-2}/N_{2p}|$ are unbounded.

Also solved by R. C. Smith and S. I. Smith and the proposer.

Annular Families

E 2892 [1981, 444]. *Proposed by F. S. Cater, Portland State University.*

Let X be a nonvoid set; let $k > 0$ be an integer. Characterize the families U of subsets (of X), card $U = 2^k$, satisfying (*) for every $A \in U$ there are exactly k sets $B \in U$ such that $(A \setminus B) \cup (B \setminus A)$ is a singleton. (Compare E 2792 [1979, 702].)

Solution by Robert Patenaude, California State University, Los Angeles. The families U are precisely those of the form $Y = \{Z: S \subseteq Z \subseteq T\}$ for some two subsets S, T of X with cardinality $(T \setminus S) = |T \setminus S| = k$. First of all it is clear that any U of the above form satisfies the conditions of the problem.

Denote, as usual, the symmetric difference $(A \setminus B) \cup (B \setminus A)$ by $A \Delta B$. If U is any family of subsets of X and $Y \subseteq X$, put

$$U^Y = \{C \Delta Y: C \in U\}.$$

It follows that $|U^Y| = |U|$ and $(U^Y)^{Y} = U$. Thus for any $Y \subseteq X$, U^Y has 2^k sets and satisfies (*) if and only if U does. Fix $Y \in U$ and put $U_i^Y = \{C \in U^Y: |C| = i\}$, so that $\varnothing = Y \Delta Y \in U^Y$ and $|U_0^Y| = 1$. For $i \geqslant 1$ each set in U_i^Y contains at most i sets in U_{i-1}^Y and thus it must be contained in at least $k - i$ sets of U_{i+1}^Y. It follows that

$$(k - i)|U_i^Y| \leqslant (i + 1)|U_{i+1}^Y|$$

and the induction gives $|U_i^Y| \geqslant \binom{k}{i}$ so

$$|U^Y| \geqslant \sum_{i \geqslant 0} |U_i^Y| \geqslant \sum_{i \geqslant 0} \binom{k}{i} = 2^k.$$

Since $|U^Y| = 2^k$ this forces each set in U_i^Y to contain exactly i sets in U_{i-1}^Y so every subset of a set in U^Y is also in U^Y. Since $|U_k^Y| = 1$, say $U_k^Y = \{D\}$, we clearly have $U^Y = 2^D$. Now, retrieve

$$U = (U^Y)^Y = (2^D)^Y = \{Z: Y \setminus D \subseteq Z \subseteq Y \cup D\},$$

so $S = Y \setminus D, T = Y \cup D$.

This problem is closely connected with E 2795 [1979, 703; 1980, 757] where (in a different formulation) it is shown that if U satisfies (*), then $|U| \geqslant 2^k$.

Also solved by the proposer.

A Totient Identity

E 2916 [1981, 763]. *Proposed by R. Sivaramakrishnan, University of Calicut, India.*

Let $\phi_2(r)$ represent the number of integers, a, $1 \leqslant a \leqslant r$, with $(a, r) = (a + 1, r) = 1$. Prove that, for $r, n \geqslant 1$, the relation

$$\Sigma \phi(nr/d^2)\, d\mu(d) = \phi(n/u)\phi(r/u)\phi_2(u)$$

holds. Summation extends over all d, $d|(n, r)$. u is the greatest common square-free unitary divisor of n and r. (A divisor d of c is called unitary if $(d, c/d) = 1$.) $\phi(\mu)$ is the Euler totient [Möbius function].

Solution by Chun-Nip Lee, student, Massachusetts Institute of Technology. The stated formula contains a misprint. A correct formula is

$$\sum \phi(nr/d^2)\, d\mu(d) = \phi(n/u)\phi(r/u)\phi_2(u^2).$$

(*Editor's note.* The greatest common square-free unitary divisor (g.c.s.f.u.d.) of p and p^α, $\alpha \geqslant 2$, is 1; this is not the same as the greatest square free unitary divisor of the greatest common divisor of two numbers.)

Proof by induction on the number of distinct prime factors of (n, r). If $(n, r) = 1$, then each side of the equation equals $\phi(n)\phi(r)$ by multiplicativity of ϕ. Now suppose that the formula is true if (n, r) contains at most k distinct prime factors, $k \geqslant 0$, and that $(n, r) = p^\eta e$, where e contains k distinct prime factors, $\eta \geqslant 1$, and $(p, e) = 1$. Write $n = p^\nu N$ and $r = p^\rho R$, with $(p, N) = (p, R) = 1$. Let U denote the g.c.s.f.u.d. of N and R and u the g.c.s.f.u.d. of n and r.

Let Σ denote the sum on the left side of the formula. We can express Σ as a sum over the divisors of e plus a sum over the remaining divisors of pe, since $\mu(d) = 0$ if d is divisible by the square of a prime. We obtain by multiplicativity

$$\begin{aligned}\sum &= \sum_{d|p^\eta e} \phi(p^{\nu+\rho}NR/d^2)\, d\mu(d)\\ &= \left(\phi(p^{\nu+\rho}) - p\phi(p^{\nu+\rho-2})\right)\sum_{d|e}\phi(NR/d^2)\, d\mu(d).\end{aligned}$$

The induction hypothesis now gives

$$\sum = \left(\phi(p^{\nu+\rho}) - p\phi(p^{\nu+\rho-2})\right)\phi\left(\frac{N}{U}\right)\phi\left(\frac{R}{U}\right)\phi_2(U^2).$$

If $\nu = \rho = 1$, then $u = pU$ and

$$\begin{aligned}\sum &= (p^2 - 2p)\phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(U^2)\\ &= \phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(u^2),\end{aligned}$$

since ϕ_2 is multiplicative and $\phi_2(p^2) = p^2 - 2p$.

If one of ν, ρ exceeds 1, then $u = U$ and

$$\begin{aligned}\sum &= \left(p^{\nu+\rho} - 2p^{\nu+\rho-1} + p^{\nu+\rho-2}\right)\phi\left(\frac{N}{U}\right)\phi\left(\frac{R}{U}\right)\phi_2(U^2)\\ &= \phi(p^\nu)\phi(p^\rho)\phi\left(\frac{N}{u}\right)\phi\left(\frac{R}{u}\right)\phi_2(u^2)\\ &= \phi\left(\frac{n}{u}\right)\phi\left(\frac{r}{u}\right)\phi_2(u^2).\end{aligned}$$

This completes the induction.

ANSWERS TO PHOTOS ON PAGE 624

Top: A. G. Kurosh (algebra, and in particular, group theory), taken in 1958. Bottom: L. Schwartz (analysis, and in particular, distributions) taken in 1969.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by March 31, 1984. The solver's full post-office address should be on each sheet.

6442. *Proposed by Vladimir Naroditsky, San Jose State University.*

Show that if

$$\lim_{t\to\infty}\sum_{k=1}^{N} a_k \exp(i\lambda_k t)$$

exists, then either all λ's are 0 or all a's are 0.

6443. *Proposed by C. Thron and B. Tomaszewski, University of Wisconsin.*

Let $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_{2n+1} \geqslant 0$ be a decreasing sequence of positive real numbers and let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{2n+1}$ be a Bernoulli sequence of independent random variables, i.e., $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = \frac{1}{2}$. Prove that

$$P\left(\left|\sum_{i=1}^{2n+1}\varepsilon_i a_i\right| < a_1\right) + \frac{1}{2}P\left(\left|\sum_{i=1}^{2n+1}\varepsilon_i a_i\right| = a_1\right) \geqslant \frac{1}{2^{2n+1}}\binom{2n+1}{n}.$$

6444. *Proposed by Fred Galvin, University of Kansas.*

A topological space X is *functionally countable* if every continuous real-valued function defined on X has a countable range. Prove or disprove: the product of two functionally countable spaces is functionally countable.

SOLUTIONS OF ADVANCED PROBLEMS

$\lim_{n\to\infty}|\sin n|^{1/n} = 1$

6379 [1982, 134]. *Proposed by Robert Curry and James O. Friel, California State University, Fullerton.*

Find $\liminf_{n\to\infty}|\sin n|^{1/n}$.

Solution by Gunnar Berg, Uppsala Universitet, Matematiska Institutionen, Sweden. Let m, n be positive integers such that $m > 1$ and $|n - m\pi| < \pi/2$. Then $m < n$ and, by a theorem of K. Mahler ["On the approximation of π", *Indag. Math.*, 15 (1953) 30–42], $|n - m\pi| > m^{-41}$. Hence

$$1 > |\sin n| = |\sin(n - m\pi)| > \frac{2}{\pi}|n - m\pi| > \frac{2}{\pi}n^{-41}.$$

It follows that $\lim_{n\to\infty}|\sin n|^{1/n} = 1$.

Also solved by Michael Golomb, Robert B. Israel (Canada), Kee-Wai Lau (Hong Kong), O. P. Lossers (The Netherlands), T. K. Louton & C. C. Rousseau, John M. Masley & John G. Milcetich, Daniel A. Rawsthorne, Joseph J. Roseman, and José Felipe Voloch (Brazil).

John M. Masley & John G. Milcetich based their solution on a result of Cijsouw (see p. 95 in *Transcendance Theory: Advances and Applications*, edited by A. Baker and D. W. Masser) which states that there is an absolute positive constant c such that if ζ is any algebraic number of degree N and height H, then

$$|\pi - \zeta| > \exp(-cN(1 + N\log N + \log H)(1 + \log N)).$$

They observed that it follows easily from this result that $\lim_{n\to\infty}|\sin n\zeta|^{1/n} = 1$ when ζ is any nonzero algebraic number.

Functions Agreeing Infinitely Often

6380 [1982, 214]. *Proposed by Juris Steprāns, University of Toronto.*

Let l be a finite ordinal, and let ${}^{\omega}l$ denote the set of functions from ω to l where ω is the first infinite ordinal. Is the proposition P: If $F \subseteq {}^{\omega}l$ and $\operatorname{card}(F) < 2^{\aleph_0}$, then there exists $g \in {}^{\omega}l$ such that for all $f \in F$ there are infinitely many m for which $f(m) = g(m)$ true, false, or independent of ZFC?

Solution by the proposer. The proposition is true. First note that it suffices to show that if $F \subseteq {}^{\omega}l$ and $\operatorname{card}(F) < 2^{\aleph_0}$, then there is $g \in {}^{\omega}l$ such that for each $f \in F$ there is at least one integer m such that $f(m) = g(m)$. To see this let $\{A_n: n \in \omega\}$ partition ω into disjoint infinite subsets. Let $F_n = \{f \restriction A_n: f \in F\}$ and apply the proposition to find $g_n: A_n \to l$ such that for each $f \in F_n$ there is at least one $m \in A_n$ such that $f(m) = g_n(m)$. Then let $g = \cup\{g_n: n \in \omega\}$. The proposition is proved as follows. Let $D(l)$ be the discrete topology on l and let $D(l)^{\mathbb{R}}$ have the usual product topology. It is known (see page 111, *General Topology*, by R. Engelking, Warsaw, 1977) that $D(l)^{\mathbb{R}}$ is separable. Hence, let $\{h_k: k \in \omega\}$ enumerate a dense subset of $D(l)^{\mathbb{R}}$.

For $r \in \mathbb{R}$ define $g_r: \omega \to l$ by $g_r(k) = h_k(r)$. Now suppose that F is a counterexample to the proposition. Then for each $r \in \mathbb{R}$ there is $f \in F$ such that for each $m \in \omega$, $f(m) \neq g_r(m)$. Since $\operatorname{card}(F) < 2^{\aleph_0}$ it follows from the pigeonhole principle that for some $f \in F$ and some $X \subseteq \mathbb{R}$ such that $\operatorname{card}(X) = l$, for each $r \in X$ and each $m \in \omega$, $f(m) \neq g_r(m)$. Now let $b: X \to l$ be a bijection, then $[b] = \{h \in D(l)^{\mathbb{R}}: b \subseteq h\}$ is an open set in $D(l)^{\mathbb{R}}$ and hence there is $k \in \omega$ such that $h_k \in [b]$ (i.e., $b \subseteq h_k$). Choose $t \in X$ such that $b(t) = f(k)$. Then $g_t(k) = h_k(t) = b(t) = f(k)$. This is a contradiction.

F. Galvin pointed out that the result has been obtained earlier by J. Baumgartner (unpublished).

Evaluation of an Integral

6382 [1982, 214]. *Proposed by Hing-kam Lam, The Chinese University of Hong Kong.*

Evaluate $\int_0^1 \ln x \ln^2(1-x)\,dx/x$.

Solution by Sidney Heller, Brookhaven National Laboratory, Upton, New York. Let

$$f(x) = \int_0^x \frac{\ln(1-t)}{t}\,dt.$$

Then

$$f(x) = -\sum_{n=1}^{\infty} \frac{x^n}{n^2} \quad \text{for } 0 \leqslant x \leqslant 1,$$

and hence

$$\begin{aligned}
\int_0^1 \frac{\ln x \ln^2(1-x)}{x}\,dx &= -\int_0^1 f(x)\left(\frac{\ln(1-x)}{x} - \frac{\ln x}{1-x}\right) dx \\
&= -\int_0^1 f(x) f'(x)\,dx - \int_0^1 \frac{\ln x}{1-x} \sum_{n=1}^{\infty} \frac{x^n}{n^2}\,dx \\
&= -\frac{1}{2} f^2(1) - \sum_{n=1}^{\infty} \left(1 + \frac{1}{2^2} + \cdots + \frac{1}{n^2}\right) \int_0^1 x^n \ln x\,dx \\
&= -\frac{1}{2}\zeta^2(2) + \sum_{n=1}^{\infty} \left(1 + \frac{1}{2^2} + \cdots + \frac{1}{n^2}\right) \frac{1}{(n+1)^2} \\
&= -\frac{1}{2}\zeta^2(2) + \sum_{1 \leqslant m < n}^{\infty} \frac{1}{m^2 n^2} \\
&= -\frac{1}{2}\zeta^2(2) + \frac{1}{2}(\zeta^2(2) - \zeta(4)) = -\frac{1}{2}\zeta(4) = -\frac{\pi^4}{180}.
\end{aligned}$$

Also solved by Paul S. Bruckman, Donald Caccia, The Chico Problem Group, C. Georghiou (Greece), M. L. Glasser, Nathaniel Grossman, A. P. Guinand (Canada), John R. Hatcher, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), T. K. Louton & C. C. Rousseau, L. E. Mattics, William A. Newcomb, Otto G. Ruehr, L. Van Hamme (Belgium), John S. White, Doug Wiedemann, Joseph Wiener, and the proposer. A partial solution was received from Murray R. Spiegel.

A number of solvers pointed out that the integral and many like it are evaluated in *Polylogarithms and Associated Functions* (North-Holland, New York, 1981), by Leonard Lewin.

A Limit Involving Euler's Totient

6383 [1982, 278]. *Proposed by Eliot Jacobson, University of Arizona, Tucson.*

Let $\phi(x)$ denote the totient function, and define recursively $\phi^r(x) = \phi(\phi^{r-1}(x))$. Let

$$A_m(n) = |\{x \leqslant n\colon \phi^k(x) = m, \text{some } k\}|.$$

Show that $\lim_{n\to\infty} A_m(n)/n$ exists for all $m > 0$, and calculate its value.

Solution by the University of South Alabama Problem Group, Mobile, Alabama. If $m = 2^s$, the limit is 1; otherwise it is zero. Indeed, if m is divisible by an odd prime and $\phi(k) = m$, then k is divisible by an odd prime and k has at most as many prime divisors (counting multiplicity) as m. So, in this case, numbers x such that $\phi^r(x) = m$ must have a bounded number of prime divisors and the set of such numbers has natural density zero. (See Niven and Zuckerman, *An Introduction to the Theory of Numbers*, 3rd ed., Theorem 11.8.) On the other hand, if $m = 2^s$ and if x has more than s distinct prime divisors, then $\phi^r(x) = m$ for some r and the set of such numbers x has natural density 1.

Also solved by the proposer. A partial solution was received from Carl Pomerance.

Numbers of the Form $m^p - n$

6384 [1982, 278]. *Proposed by Barry Powell, Kirkland, Washington.*

(a) For any positive integers m and n for which $m - n \neq 1$, prove that the set $\{m^p - n\}$, (where $p = 2, 3, 5, 7, \ldots$ is the set of primes) contains infinitely many composites.

(b)* Give an infinite set of n such that $(n+1)^p - n$ is composite infinitely often ($p = 2, 3, 5, \ldots$).

Solution by Robert B. Israel, University of British Columbia, Vancouver, B.C., Canada. We must of course assume that $m \neq 1$. Instead of excluding the case $m - n = 1$, we exclude only the case $m = 2$, $n = 1$ (which yields the Mersenne numbers). Suppose therefore that $m > 1$, $mn > 2$ and that q is a prime factor of $mn - 1$. If $p = k(q-1) - 1$, where k is a positive integer, then $m(m^p - n) \equiv 1 - mn \equiv 0 \pmod q$ and so $m^p - n$ is divisible by q. By Dirichlet's theorem, there are infinitely many primes p of this form. Hence $m^p - n$ is composite for infinitely many primes p when $m > 1$, $mn > 2$. In particular, $(n+1)^p - n$ is composite for infinitely many primes p whenever $n > 1$. This deals with both parts of the problem.

Also solved by Miroslav D. Ašić (Yugoslavia), Mihály Bencze (Romania), Robert Breusch, Hugh M. Edgar, F. Göbel (The Netherlands), E. Grosswald, Mark Kantrowitz, Keith A. Kearnes, O. P. Lossers (The Netherlands), Aaron Meyerowitz, Peter Schumer, Robert E. Shafer, H. Shank (Canada), Claudia Spiro, Richard Stong, University of South Alabama Problem Group, Keith Wayland (Puerto Rico), and the proposer.

A number of the solvers failed to consider the case $m - n = -1$.

A Simple Closed Curve Circumscribing No Rectangle

6385 [1982, 279]. *Proposed by Mark D. Meyerson, U.S. Naval Academy, Annapolis, MD.*

Prove or disprove: Every simple closed curve in Euclidean space contains the vertices of a rectangle. (It is known to be true in the Euclidean plane.)

Solution by Miroslav D. Ašić, University of Belgrade, Yugoslavia. The assertion is false. Let L be the curve in $\mathbb{R}^3$ consisting of four line segments AB, BC, CD, DA where $A = (0,0,0)$, $B = (0,0,1)$, $C = (1,0,0)$, $D = (1,1,0)$. Assume there is a rectangle $PQRS$ with vertices on L. Clearly, no two vertices of the rectangle can lie on the same (closed) line segment. This means that each (open) line segment contains exactly one of the points P, Q, R, S. It is easily seen that the consecutive vertices of the rectangle must lie on consecutive line segments of L; without loss of generality we can assume that P, Q, R, S are on AB, BC, CD, DA respectively. But then PQ and QR are parallel to AC and BD respectively, which is impossible since AC and BD are not orthogonal.

Also solved by Benny N. Cheng, Robert B. Israel (Canada), P. B. Kronheimer (England), O. P. Lossers (The Netherlands), J. G. Mauldon, Marion W. Orlowski (South Africa), M. Pachter (South Africa), J. Schaer (Canada), Allen J. Schwenk (Canada), Paulo Ney Rocha de Souza (Brazil), and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Infinite Processes: Background to Analysis. By A. Gardiner. Springer-Verlag, New York, 1982. x + 306 pp.

R. P. BOAS
Department of Mathematics, Northwestern University, Evanston, IL 60201

Once upon a time there were an Algebraist and an Analyst who were participating in Hermann Weyl's current-literature seminar. The Algebraist undertook to present a celebrated paper on the distribution of the values of meromorphic functions. After a while, being unable to verify the transition from a certain sentence to the next one, the Algebraist appealed to the Analyst for help, and was provided with a six-page argument for that particular point. The Algebraist, aghast, complained, "But in Algebra we *prove* the theorems."

This story brings out the wide individual variation in what is accepted as "rigorous mathematics." In this review, I am going to substitute the term "careful mathematics" (since standards of rigor are so variable). We should distinguish careful mathematics from heuristic mathematics, which is (mistakenly) sometimes considered not to be mathematics at all. All mathematicians use heuristic mathematics some of the time, but all serious mathematicians use careful mathematics a substantial part of the time.

For students who hope to become serious mathematicians, there are courses in which they are introduced to careful proofs, often for the first time in their lives (now that careful Euclidean geometry is no longer a staple of the high school curriculum). As Gardiner remarks (p. 270), the actual writing out in detail of the formal proofs is tedious; "what is fun is discovering the answer." This is true, but it does not alter the fact that today's college students (in the USA, anyway) are not much accustomed to careful proofs.

If there are courses, there will be textbooks. What should these textbooks be like? Gardiner's

answer is embodied in this book, which he suggests is useable either before, during, or after the first course in careful mathematics. If you are going to have to persuade students that they need a more precise understanding of the mathematics they thought they knew, and if they have little experience with careful mathematics, then you must first persuade them that there is something wrong with what they thought they knew; for only then will the new ideas have a chance of becoming credible. Gardiner seems to believe that this persuasion is best performed by means of history, an approach that should appeal to those students who share the current preoccupation with "roots." Advocates of the historical approach (but not Gardiner) often quote Santayana, "Those who cannot remember the past are condemned to repeat it." [**1**] This rings true enough for statesmen, but the history of mathematics is largely the study of the mistakes of our predecessors, which we no more need to repeat than we need to study the abacus after having learned to use a computer. Most mathematicians find it more profitable to concentrate on what is correct (or presumed to be correct) right now. It is entertaining to learn what Leibniz and John Bernoulli thought about $\log(-1)$ (p. 262—without references!), but this does not particularly illuminate modern complex analysis. Biologists say that ontogeny recapitulates phylogeny, that is, the development of the individual parallels the development of the species; but no such principle seems to be valid for mathematics.

In Gardiner's words, "the whole book is essentially a study of why one must and how one can introduce *precision* into certain infinite processes..." (p. 4). Chapter I is introductory; Chapter II deals with numbers, in great detail, from the positive integers to the irrationals. Chapter III analyzes the ideas of area and volume and also of the length of curves. Chapter III is considerably shorter than Chapter II, but the subject has a much shorter history. Both chapters are lucid and extremely detailed; so much so that the book gives the impression of "a missionary talking to cannibals" (Littlewood's phrase [**2**]). American readers need to be alert for occasional Anglicisms, however; for example, the phonetic spelling of the name of the letter τ as "tor," final *r* being silent in standard British pronunciation (p. 148).

Chapter 4, on functions, traces the history of "function" from its beginnings down to the now conventional informal formal definition. The remoter past is fully dealt with, but there is no mention of the many alternative ways of thinking about functions, for instance as operators, mappings, random variables, fields (in the physicists' sense), or even as sets of ordered pairs. I would have thought that experience with other interpretations of the word would be more helpful than examples of the tentative gropings of the past. It does seem that many of the difficulties associated with "function" could have been avoided if earlier generations had had "the vision to give it existence through definition." [**3**]

Now, I am well aware that some people never succeed in grasping the ideas involved in that first contact with careful mathematics; and that some others approach the subject with a wariness more appropriate to a subject like spherical trigonometry: they can do the problems but they can't see the point. "What can rigorous proofs be used for?" they complain. Gardiner seems to think that all students are like these people. He writes (p. v), "I have consistently resisted the temptation to believe that the process of understanding these ideas can be magically accelerated by confronting the reader with ready-made abstractions."

There is, however, a class of students for whom abstractions can be liberating; and my own experience persuades me that this class is not empty. Through my sophomore year in college, the closest I came to knowing anything about careful mathematics was hearing J. L. Doob (then a graduate student) holding forth on the lack of rigor in Osgood's *Calculus*. The next year I took Walsh's course in complex analysis; this was quite a careful course (in the sense in which I am using "careful"), and sufficiently abstract so that my year's lecture notes contain no pictures at all (my idea, not Walsh's). This course was a revelation: to say that I fell in love with the subject would be quite an accurate description of my reaction. I can speak with assurance only of myself, but I have known a fair number of professional mathematicians who started out in much the same way. On the whole, we *like* abstractions, whether ready-made or not. Most of us, however, try to

present them (to students) with introductory examples and illustrations of their applicability. In short, we try to make new ideas credible by embedding them in more familiar mathematical settings, rather than as the culmination of the ideas of the past.

References

1. G. Santayana, The Life of Reason, 1905, quoted in A. L. Mackay, The Harvest of a Quiet Eye, 1977.
2. J. E. Littlewood, quoted in the preface to the 7th edition of G. H. Hardy's A Course of Pure Mathematics.
3. W. F. Osgood, Advanced Calculus, 1925, p. 339, where this phrase was used in a different context.

LETTERS TO THE EDITOR

Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

Nowadays, the usual derivation of canonical forms for $n \times n$ matrices A over a field F involves such matters as invariant subspaces and cyclic vectors. Briefly, one constructs from a given matrix A a certain $F[x]$-module V. Since $F[x]$ is a principal ideal domain, there is a very good analogy between $F[x]$-modules and abelian groups (Z-modules). In particular, the analog of V is a finite abelian group. Just as a finite abelian group G is characterized by some integers (G is a direct sum of cyclic groups and we may take the orders of the summands), so the module V is characterized by some polynomials (the invariant factors of A). I like this proof; the analogy is quite beautiful, and one can understand what is going on.

But current proofs have a defect; given a matrix A, they do not indicate how to compute its invariant factors. (Note that the minimal polynomial of A is one of them.) Thus, there are two articles in the January 1983 issue of the MONTHLY that seem to have overlooked an old theorem. ("An algorithmic derivation of the Jordan canonical form," by Fletcher and Sorensen, and "An algorithm for the minimal polynomial of a matrix," by Gelbaum.) The theorem says that if B is a matrix with (polynomial) entries in $F[x]$, then one can put B in diagonal form $\operatorname{diag}(g_1(x), g_2(x), \ldots, g_n(x))$, where $g_i(x) | g_{i+1}(x)$, using elementary row and column operations. (In so doing, one needs the Euclidean algorithm for the gcd of two polynomials.) In particular, this can be done for $B = xI - A$. The nonconstant $g_i(x)$ are the invariant factors of A, and $g_n(x)$ is the minimal polynomial of A. A purely matrix-theoretic proof of this result can be found, for example, in "Introduction to Algebraic Theories," by A. A. Albert, University of Chicago Press, 1941.

Joseph Rotman
Department of Mathematics
University of Illinois
Urbana, Illinois 61801

MISCELLANEA

114. There exists a passion for understanding just as there exists a passion for music. This passion is more common in children, and is usually lost with age. Without this passion there would be neither mathematics nor natural science. This passion which has always existed in me has never lost its sparkle.

—*Albert Einstein: The Man and his Theories*, Souvenir Press, 1961, p. 114.

INDEX TO VOLUME 90, 1983
THE AMERICAN MATHEMATICAL MONTHLY

AUTHOR INDEX

INDEX OF SPECIAL RECURRING TOPICS

SUBJECT INDEX

This index uses the current version of AMS 1980 Mathematics Subject Classification.

PROBLEMS AND SOLUTIONS

PROBLEMS PROPOSED

PROBLEMS SOLVED

SOLUTIONS

Numbers in boldface type refer to problems; those in lightface, to pages.

REVIEWS

Names of authors are in ordinary type; those of reviewers, in capitals.

LETTERS TO THE EDITOR

MISCELLANEA

ERRATA AND ADDENDA

Referring to his joint paper with Charles Blair, "A universal entire function," which appeared in this MONTHLY, 5 (1983) 331–332, Lee A. Rubel writes as follows: "Robert Burckel has kindly drawn our attention to the paper by Gerald MacLane, 'Sequences of derivatives and normal families,' J. Analyse Math., 2 (1952) 72–87 [MR 14, p. 74] [Zbl. 49, p. 57]. Among other things, MacLane constructed 'a ubiquitous entire function' by the same method we use, to have the same 'universal' property. However, he also gets estimates on the rate of growth of the function, so that his proof is somewhat more complicated than ours."

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 90, Number 10 December 1983

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. This is necessary for indexing purposes. Submit material to the editor or appropriate associate editor as indicated below.

EDITOR

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

AUTHOR VS. REFEREE: A CASE HISTORY FOR MIDDLE LEVEL MATHEMATICIANS

ROBERT C. THOMPSON

Mathematics Department, University of California, Santa Barbara, CA 93106

This note evolved from a referee's rejection of a research paper that I wrote. The reasoning behind the rejection was perhaps unusual, and leads to a not altogether trivial question concerning the role of the referee in the professional development of a mathematician. The discussion will be more candid than is customary, and this may add spice to the article, since confession of failure, or even of sin, is always interesting.

The first section describes the mathematical problem that was investigated, and the second outlines what was proved concerning it and what was not. A research paper was written, and the motivation to publish it is given in the third section. The resulting referee's report is described in the fourth section, and in the fifth the somewhat psychological issues it leads to are examined. Those readers who submit papers to research journals may wish to reflect on these issues in the light of their own experience. The sixth section briefly discusses waiting times, and a few concluding remarks are in the seventh.

1. The Mathematical Question. Consider $n \times n$ matrices having integer entries. If A is such a matrix, it is known that there exist unimodular matrices U and V of integers such that

$$UAV = \begin{bmatrix} a_1 & & & & \\ & a_2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & \cdot & \\ & & & & a_n \end{bmatrix}$$

is a diagonal matrix with the diagonal elements forming a divisibility chain:

$$a_1|a_2| \cdots |a_n.$$

(The symbol | signifies divides, with the convention that $0|a$ means that $a = 0$. To say that U is unimodular means that $\det U = \pm 1$.) This is the diagonal form published by H. J. S. Smith in 1861 and now universally known as "the Smith canonical form." The diagonal entries $a_1, \ldots, a_n$ are called the invariant factors of A; they are unique to within a plus or minus sign. In the sequel $a_1, \ldots, a_{n-1}$ will always be taken nonnegative and a_n will be chosen so that

$$\det A = a_1 a_2 \cdots a_n.$$

Let B and C also be $n \times n$ matrices of integers, with invariant factors

$$b_1| \cdots |b_n \text{ for } B, \qquad b_1 \cdots b_n = \det B,$$
$$c_1| \cdots |c_n \text{ for } C, \qquad c_1 \cdots c_n = \det C.$$

The issue to be addressed is how invariant factors behave when matrices add. This question is significant because a great deal is known about the behavior of invariant factors under matrix multiplication, whereas essentially nothing is known about their behavior under addition. Write

Robert C. Thompson: I received my B.A. and M.A. degrees from the University of British Columbia and a Ph.D. from the California Institute of Technology. The thesis supervisors were Marvin Marcus (for the M.A.) and Olga Taussky Todd (for the Ph.D.) Most of my postdoctorate academic life has been at the University of California, Santa Barbara. My research program is concentrated in linear algebra, usually involving spaces over fields but sometimes over rings. I greatly enjoy teaching anything mathematical, and especially if there is an opportunity to present the pure mathematician's outlook to an applied audience.

$C = A + B$; what does this imply about the a_i, b_i, c_i?

Since a_1 is the greatest common divisor of the elements of A (similarly for b_1 and B, c_1 and C), an easy argument shows that any common factor of a_1 and b_1 is also a factor of every a_i, b_i, c_i and of each element of A, B, C. It may therefore be cancelled. Thus generality will not be lost by taking a_1, b_1, c_1 pairwise relatively prime. This simplifying assumption will henceforth hold. To avoid trivial cases, take a_1, b_1, c_1 nonzero.

By Smith's theorem, there always exist unimodular matrices U_1, U_2, V_1, V_2 such that

$$U_1AV_1 = \begin{bmatrix} a_1 & & & \\ & \cdot & & 0 \\ & & \cdot & \\ 0 & & & \cdot \\ & & & a_n \end{bmatrix}, \quad U_2BV_2 = \begin{bmatrix} b_1 & & & \\ & \cdot & & 0 \\ & & \cdot & \\ 0 & & & \cdot \\ & & & b_n \end{bmatrix}.$$

A consequence of this is the following equation in $2n \times 2n$ matrices:

$$\begin{bmatrix} C & \cdot \\ \cdot & \cdot \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} a_1 & & & & & \\ & \ddots & & & 0 & \\ & & a_n & & & \\ & & & b_1 & & \\ & 0 & & & \ddots & \\ & & & & & b_n \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ V_2 & I \end{bmatrix}.$$

Since the first and last matrices on the right are unimodular, this equation reveals that C is an $n \times n$ submatrix of a $2n \times 2n$ matrix having the same invariant factors as the diagonal matrix $D = \mathrm{diag}(a_1,\ldots,a_n, b_1,\ldots,b_n)$. Two facts will now be brought to bear:

(i) It is relatively easy, though not completely trivial, to prove that the invariant factors $d_1 | \cdots | d_{2n}$ of D satisfy

$$(a_i, b_j) | d_{i+j-1},$$

where $(\,,)$ signifies the greatest common divisor. The proof is in **[6]**.

(ii) It is a somewhat standard fact, though not too well known, that the invariant factors $c_1 | \cdots | c_n$ of C and those $d_1 | \cdots | d_{2n}$ of the larger matrix containing C satisfy

$$d_i | c_i \quad \text{for } 1 \leqslant i \leqslant n.$$

These facts imply that

$$(a_i, b_j) | c_{i+j-1} \tag{1}$$

for all indices i, j for which the subscripts lie in the range $1, n$.

The inequalities (= divisibility relations) visible in (1) are very much like a rather well-known family of inequalities for the eigenvalues of a sum or product of complex matrices. Example: If $A + B = C$, with A, B, C Hermitian having eigenvalues $\alpha_1 \geqslant \cdots \geqslant \alpha_n$, $\beta_1 \geqslant \cdots \geqslant \beta_n$, $\gamma_1 \geqslant \cdots \geqslant \gamma_n$, respectively, then $\gamma_{i+j-1} \leqslant \alpha_i + \beta_j$. This lovely inequality was found by H. Weyl in 1912 **[9]**. Extensions of it involving scatterings of terms are now known, the simplest being **[5]**

$$\sum_{s=1}^{m} \gamma_{i_s+j_s-s} \leqslant \sum_{s=1}^{m} \alpha_{i_s} + \sum_{s=1}^{m} \beta_{j_s} \quad \text{if } i_1 < \cdots < i_m,\ j_1 < \cdots < j_m. \tag{2}$$

The indices on the terms here belong to a family of sequences occurring in the representation theory of the symmetric group and first found by D. E. Littlewood and A. R. Richardson in 1934 **[2]**. An immediate question is whether extensions of (1) along the lines of (2) are valid for the

invariant factors of integer matrices. It turned out that most conjectured extensions of (1) were almost instantly seen to be wrong, and those that weren't could always be deduced from (1). Thus it appears that no further conditions like (2) have to be considered when studying the invariant factors of a sum of integer matrices.

But there is an obvious further condition of quite a different form. Modulo a_1, the matrix A becomes zero, so that

$$\det B \equiv \det C (\mathrm{mod}\ a_1).$$

Observing now that everything so far said may be applied twice more by rewriting $A + B = C$ as $A + (-C) = -B$ and as $B + (-C) = -A$, we obtain the following theorem.

THEOREM 1. *The invariant factors of a sum $A + B = C$ of integral matrices satisfy*

$$(a_i, b_j) | c_{i+j-1}, \qquad (a_i, c_j) | b_{i+j-1}, \qquad (b_i, c_j) | a_{i+j-1},$$

$$\det A \equiv (-1)^n \det B (\mathrm{mod}\ c_1), \quad \det A \equiv \det C (\mathrm{mod}\ b_1), \quad \det B \equiv \det C (\mathrm{mod}\ a_1).$$

Question. Are the conditions of Theorem 1 sufficient?

That is, if integers $a_1 | \cdots | a_n$, $b_1 | \cdots | b_n$, $c_1 | \cdots | c_n$ are given satisfying the conditions of Theorem 1, can integer matrices A, B, $C = A + B$ be found having the a_i, b_i, c_i as their respective invariant factors? This was the mathematical problem that I addressed in the research paper mentioned earlier.

2. Partial Answers. If a_i, b_i, c_i are integers satisfying the conditions of Theorem 1, can integer matrices $A + B = C$ be found with these integers as invariant factors? The matrices are to be $n \times n$; the following discussion treats each value of n in turn. Recall the assumption made earlier that a_1, b_1, c_1 are pairwise relatively prime.

For $n = 1$, the question has a *negative* answer: even though a_1, b_1, c_1 satisfy the conditions, they need not satisfy $a_1 + b_1 = c_1$. This discouraging result suggests that a reasonably clean theorem is unlikely.

For $n = 2$, however, the answer is *positive*: If $a_1 | a_2$, $b_1 | b_2$, $c_1 | c_2$ satisfy the conditions, it is possible to construct 2×2 matrices A, B, $C = A + B$ of integers having the a_i, b_i, c_i as invariant factors. This encouraging result suggests that a reasonably clean theorem can probably be found.

The proof in the 2×2 case is not too hard, but not obvious either. Here are a few details. Since a_1, b_1 are relatively prime, there are integers r, s such that $ra_1 + sb_1 = 1$. Write

$$A = a_1 \begin{bmatrix} rc_1 & b_1 \\ b_1 z_{21} & rc_2 + b_1 z_{22} \end{bmatrix}, \quad B = b_1 \begin{bmatrix} sc_1 & -a_1 \\ -a_1 z_{21} & sc_2 - a_1 z_{22} \end{bmatrix}, \quad C = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix},$$

where z_{21} and z_{22} are integers to be chosen so that A has a_1, a_2 as its invariant factors and B has b_1, b_2. It is almost immediate that only the determinantal constraints $\det A = a_1 a_2$ and $\det B = b_1 b_2$ have to be met. Using the conditions from Theorem 1, this can be shown to be possible, after a not too long but not altogether straightforward analysis. There are nine conditions, but only three are actually needed, since these three imply the remaining six.

For $n = 3$, the answer is again positive, but the proof is much, much longer. There are eighteen conditions, and all eighteen are needed, some repeatedly. The technique is similar to that used in the 2×2 case. One takes $C = \operatorname{diag}(c_1, c_2, c_3)$ to be diagonal,

$$A = a_1 \begin{bmatrix} rc_1 & b_1 & b_1 z_{13} \\ b_1 z_{21} & rc_2 + b_1 z_{22} & b_1 z_{23} \\ b_1 z_{31} & b_1 z_{32} & rc_3 + b_1 z_{33} \end{bmatrix},$$

$$B = b_1 \begin{bmatrix} sc_1 & -a_1 & -a_1 z_{13} \\ -a_1 z_{21} & sc_2 - a_1 z_{22} & -a_1 z_{23} \\ -a_1 z_{31} & -a_1 z_{32} & sc_3 - a_1 z_{33} \end{bmatrix},$$

where $z_{13}, z_{21}, z_{22}, z_{23}, z_{31}, z_{32}, z_{33}$ are unknown integers to be chosen so that A has a_1, a_2, a_3 as invariant factors (a_1 is automatic) and B has b_1, b_2, b_3. The task is a difficult diophantine problem, and it is not obvious how to proceed. Nevertheless, I found a way of proceeding, in a proof that runs for about eleven typed pages. I believe this proof to be a reasonably substantial achievement.

I made some half-hearted attempts to settle the case $n = 4$, and I failed. I was, frankly, tired of the problem after the arduous work needed to settle the 3×3 case. So, sufficiency of the conditions of Theorem 1 is *not* established for $n \geqslant 4$.

A comment is in order concerning the cases $n = 2$ and $n = 3$. The proofs just described failed to reveal any conceptual framework within which to set the problem, and so reveal few genuine clues showing how to proceed to higher dimensions. The arguments are clever, but perhaps only ad hoc. The lack of a conceptual framework is very distressing. What is the structural foundation? None is known to me.

3. The Impetus Toward Publication. At this point, I came under review for a salary increase at my university, and to justify the increase it was of course necessary to exhibit a lot of new papers. (Unfortunately, quality is hard to establish, thus quantity counts in these matters.) I got busy and wrote up a dozen papers, selecting the dozen items easiest to write up from a store of unpublished results. The invariant factor problem was one of the dozen, and was written up as a manuscript entitled *Sums of integral matrices*. Because of the great difficulty I had experienced constructing the 3×3 sufficiency proof, I was absolutely confident that the paper would be accepted without hesitation. It was submitted to the Rocky Mountain Journal of Mathematics.

4. The Referee's Report. The referee's report was short, and its contents were quite unexpected. Here it is (with a journal reference deleted):

> I suggest not publishing the enclosed paper of R. C. Thompson in the Rocky Mountain Journal of Mathematics.
>
> The problem he is looking at is: To what extent do the invariant factors of integral matrices A and B determine those of $A + B$? The problem is moderately interesting. But he hasn't solved it: He has only done the 2 by 2 and 3 by 3 cases. The computation that decides these rather special cases is a real tour-de-force, but doesn't seem to uncover any interesting ideas.
>
> Thompson has done some very interesting work on integral matrices (e.g., his paper on interlacing inequalities for invariant factors). Basically, I feel someone as good as him should not try to publish such a partial, unilluminating result.

Evidently the referee read the paper in detail, and understood it thoroughly. I interpret his report in this way: "Thompson is capable of solving the full problem, and he should get busy and do so!"

The next section, which should be regarded as the heart of this note, will discuss the referee's report from various points of view. Everything so far has merely been to set the background for the discussion. How do you respond to the issues about to be raised?

5. Discussion.

Disclaimer. Both the referee and I may well have missed the point: there may be a more or less easy solution for the invariant factor problem. The reader is challenged to find one. If one is found, the following discussion may lose its punch, but this note will still have merit since it provided the stimulus for the reader's creative effort.

To continue, however, assume that no easy solution is possible, so that the results in hand are a nontrivial partial solution to a difficult problem.

Criteria for publishability: are there any generally accepted ones? They would appear to be enough depth or breadth, plus a write-up possessing sufficient clarity and polish. Originality is a key item, and great originality will offset a presentation lacking clarity or polish.

Responsibilities of an author: what are they? They are to ensure that sufficient originality is

present, that the manuscript is technically accurate, that it is carefully prepared, and that incomplete partial results are not offered when more can be achieved with a not unreasonable additional effort.

A side comment: Every mathematician serving in an editorial capacity soon learns that many prospective authors fail to discharge these responsibilities adequately; some fail grossly. Premature submission of partial results is a common sin, and inadequate attention to the technical aspects is another. (Are the proofs correct? Has the manuscript been properly proofread? Is the grammar sound?) Unreadable manuscripts are all too common. Some writers seem to feel that referees are servants, to take care of details that were overlooked. No judgment could be more mistaken.

Responsibilities of a referee: *what are they*? They are to certify that the generally accepted standards are met, to suggest improvements, and to act as a stimulus to cajole an author into a further creative effort. A referee should be able to recognize the worth of mathematical styles and attitudes differing from his own. No referee, however, has to be an author's servant, correcting shortcomings caused by carelessness, immediate rejection being the proper action in such cases. If, however, the referee is overworked, it is sound procedure to rely on an author's reputation (if any) for quality and accuracy. This does not mean that an established mathematician should expect publication of his second class results. And every referee should be willing to give a helping hand to a novice writer.

(The harsh sentences in the two paragraphs above are caused by this author's experience as a journal editor.)

Consider now the referee's reports in the light of the headings above. Because my manuscript was prepared with *extreme* care, the quality of the presentation and the technical aspects were probably adequate. I shall assume this to be the case since the referee made no comment to suggest otherwise. I now pose, and partially answer, five somewhat psychological questions.

Question 1. Should a nontrivial partial solution of a difficult problem be published?

Many mathematicians would say "yes," but some probably would say "no." What is your opinion? The referee seemed to feel that the answer is "no," but perhaps he was meeting his duty of cajoling the author into a greater creative effort. In general, I feel that a partial solution should be published if lack of publication would risk loss of a nontrivial piece of work. The whole purpose of our professional existence is to get problems solved, and if a partial solution will contribute significantly to an ultimate complete solution that may happen many years later, then the permanent record created by publication is justified.

(Rule out publishing simply to create a long publication list: that's not unknown, and probably unprofessional, even though sometimes dictated by university promotion policies. I have at times been guilty of the sin. Are you also guilty?)

Question 2. Should the lack of a structural basis affect the merit of the results so far obtained?

The referee appeared to feel that this was the case. But mathematics is full of examples of difficult questions for which no structural framework has been found. Does this lack render the questions less interesting? Many mathematicians, those who are problem solvers rather than theory developers, would say that the lack of a structural setting adds to the appeal. The absolutely pure battle between mathematician and nature, without the corrupting influence of a lot of distracting structure, is surely the highest form of intellectual activity. So the referee appears to be on shaky ground with this aspect of his analysis. It is true that the proper conceptual framework might make the problem solvable, or at least accessible (as is the case for the question of how invariant factors behave when matrices multiply, a question leading to Young tableaux and Littlewood-Richardson sequences), but it might also reveal it to be difficult. What is your opinion on this issue?

Question 3. Was the attempted publication premature since only low dimensional cases were solved?

The referee probably felt that it was, and it might have been. However, I was willing to publish a partial result and abandon the invariant factor problem because I was attracted to some other equally difficult problems—see below.

Do you sometimes abandon a problem with only a partial solution because something else becomes more interesting, or because you feel unable to achieve more? If so, what do you do with your partial result?

(One response, which I quite deliberately adopted in another case when three years of intermittent effort yielded only a partial result, is to submit a manuscript containing conjectures and those facts that can be proved, hoping not for an acceptance but for a referee's report containing a significant idea. It *is* permissible to use the referee's talents, provided his contributions are acknowledged. If he cannot supply an idea, a case for publication may already be established. Can you guess how this gambit worked for me? Answer below.*)

Question 4. A referee might suggest that the author solve a problem in full, but the author has other problems that he wishes to attack, problems that he finds more attractive. Which course of action should the author follow?

This question could be rephrased as: "How much influence should a referee have in the professional development of a mathematician?" In the present instance, the dilemma is partly attributable to my formulation of the following conjecture, a conjecture that is probably deep and certainly very structural.

CONJECTURE. *If A and B are Hermitian matrices, there will always exist unitary matrices U and V so that*

$$e^{iA}e^{iB} = e^{i(UAU^{-1}+VBV^{-1})}, \tag{3}$$

where, of course, $i = \sqrt{-1}$.

The conjecture belongs to the interface between Lie groups and Lie algebras and can be formulated in Lie-theoretic terms. I have many partial results concerning it. The Campbell-Baker-Hausdorff formula $e^xe^y = e^z$ (where z is a formal series in commutators of x and y) of course plays a role. The conjecture can also be formulated for non-Hermitian matrices A, B, but then it can only hold for A, B in a neighborhood of 0. I have not published my partial results on this class of questions because I hope (perhaps unrealistically!) to produce a complete solution. But one thing is clear: if I work more on conjecture (3), I will have to commit all my resources to it, and cannot work on the invariant factor problem. It's one or the other, not both, since I am (unfortunately) only an average mathematican, far, far below the leaders. Should I be influenced by the referee's wishes that the invariant factor problem be solved in full?

(It is possible that a reader of this article who is expert in Lie theory may resolve conjecture (3). If that happens, I shall return to the invariant factor problem, unless it too is settled by a reader!)

Question 5. Will an author lose prestige in the eyes of a referee if a rejected paper is submitted to another journal without change? If so, should the author be influenced by this fact?

That's a tough issue, the possible loss of prestige being the only reason that I have not resubmitted *Sums of integral matrices*. In fact, I have an invitation from another journal to publish it there, in its present form, and if this were not the case, I could publish it in the journal I edit, Linear and Multilinear Algebra. But I have done nothing. What would you do in this situation?

Answer*. The paper, [8**], contained a conjecture having every appearance of being solvable in a few lines. The referee took 14 months to prepare a very accurate report, recommending acceptance, but not furnishing the hoped for idea. The journal's editor, noting the referee's observation that the conjecture ought to have a short solution, chose to overrule the recommendation and rejected the paper. It remains unpublished and the conjecture unsolved. Clearly, the editor may be a major factor in the author-referee configuration.

The note you are presently reading—*Author vs. referee*—is substantially based on a lecture that I gave at the matrix conference run by S. Pierce at the University of Toronto, August, 1982. Some members of the audience felt that the paper should immediately be resubmitted, but others could understand my reluctance to do so.

There may be other questions that arise naturally. Do any occur to you? Here is one: The world of mathematics is populated mainly by average mathematicians, since the very best of us are so few. Journals obviously cater to the very best. To what extent should they cater to the average? One may argue that civilized society needs mathematicians, and that the middle level mathematician undoubtedly is a better contributor to society if journals exist that will publish his middle level theorems. What is your opinion?

6. Waiting Times. How long should an author wait before expecting to receive a referee's report? Many authors expect a report within a few months, and some expect one sooner. A clearly written paper almost always gets a report more quickly than a badly written one, and to this extent an author can do a lot to help his own case. My experience with papers that I have submitted to journals is: minimum waiting time until an editorial decision is reached—a few days, maximum-four years, typically—a few months. The four year wait led to an acceptance, probably attributable to the referee's sense of guilt. To this extent, it may pay to hope for a referee who is not speedy.

A fair rule of thumb seems to be. A referee is entitled to six months in which to prepare his report, but not longer. This means: an author should not start to complain until six months have elapsed. The six month value is somewhat arbitrarily chosen. What do you think the correct figure should be?

Some mathematicians are very conscientious about meeting their obligations as referees. Others are extremely remiss. It really seems unfair that the latter cannot be somehow penalized. Is there an imaginative journal editor somewhere who can invent the appropriate punishment?

How long did the referee of *Sums of integral matrices* take? About six weeks, an excellent performance in view of the difficult 3×3 proof.

7. Some Comments.

(i) The invariant factor theorems attributed to me in the referee's report were also obtained by E. Marquis de Sa [**3**], and certain parts of them can be deduced from facts in Bourbaki. The referee should have known that! (Perhaps he did.)

(ii) The question of the similarity invariant factors of a sum of matrices with entries in a field has been studied in [**1**] with further results in the doctoral thesis of E. Marquis de Sa.

(iii) Did I get my salary increase? I did, and fortunately it was granted before the rejection was received, making it unnecessary to jeopardize the increase by having to remove an item from my publication list. (Are you as lucky?)

(iv) Of the dozen papers mentioned above, how many were accepted? Obviously not all, and the exact number of acceptances will be left for the reader to guess. It is true, though, that some of the accepted papers were weaker than the rejected one discussed in this article. This seems to imply that not all referees adhere to the same standards. Should an author resubmit a rejected paper, hoping for an "easier" referee? The tactic is surely not unknown. Does it conform to established standards of professional conduct? Probably not. A much better ploy is to make the paper stronger, revising as necessary, before resubmitting. However, an author is sometimes (not always!) a better judge of the mathematical situation than the referee: If the contents of a rejected paper form a step in a well conceived campaign toward a substantial objective, resubmission without change may be justified.

The lack of a universal standard for evaluating mathematical papers was visible in the opinions expressed by the four referees of the article before you, *Author vs. referee*. The report given me by the MONTHLY's editor suggests that of the positions taken by the four about *Sums of integral matrices*, one supported me, one supported the referee, one could see both sides of the issue, and

the position of the fourth was not disclosed. So much for unanimity! The four did not agree on the merit of *Author vs. referee* either.

Sometimes a referee's judgement is later seen to be wrong. Now and then I am sent manuscripts to appraise, and I do reject some. Two that I rejected were later published without change in other journals, and in retrospect I think both authors took the correct path. Mistakes do occur!

(v) It is to be emphasized that I bear no grudge against the referee for rejecting my paper. After many years as a professional mathematician, I have learned to accept such twists of fate stoically. A major point is that the rejection has created uncertainty in my mind: should I continue with the invariant factor problem, or abandon it for pastures that I presently perceive to be more attractive? (Can you be similarly unmoved when one of your papers is rejected? The correct, professional, response following a rejection is to examine objectively the reasons behind it.)

(vi) No statistics seem to be available on journal rejection rates. An informed guess is that most journals reject roughly 40% of submitted manuscripts, and some (including the Monthly) much more. Do you think too many manuscripts are rejected? Or too few?

(vii) I believe this article to be in good taste, even though rather candid, addressing an annoying fact of life with which most middle level mathematicians must contend. Do you agree? Or should it have been rejected? Readers may send comments to me.

Summary. The interface between author and referee is an uneasy one.

References

1. G. N. de Oliveira, E. Marques de Sa, and J. A. Dias da Silva, On the eigenvalues of $A + XBX^{-1}$, Linear and Multilinear Algebra, 5 (1977) 119–128.

2. D. E. Littlewood and A. R. Richardson, Group characters and algebra, Philos. Trans. Roy. Soc. London Ser. A, 233 (1934) 99–141.

3. E. Marques de Sa, Imbedding conditions for λ matrices, Linear Algebra Appl., 24 (1979) 33–50.

4. H. J. S. Smith, On systems of linear indeterminate equations and congruences, Philos. Trans. Roy. Soc. London Ser. A, 151 (1861) 293–326; Collected Works I, 367–409.

5. R. C. Thompson (with L. J. Freede), On the eigenvalues of a sum of Hermitian matrices, Linear Algebra Appl., 4 (1971) 369–376.

6. R. C. Thompson, The Smith invariants of a matrix sum, Proc. Amer. Math. Soc., 78 (1980) 162–165.

7. R. C. Thompson, Sums of integral matrices, manuscript.

8. R. C. Thompson, p-adic matrix valued inequalities, manuscript.

9. H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen, Math. Ann., 71 (1912) 441–479.

MISCELLANEA

115.

To see ourselves as others see us?

> "If you try to understand fibre bundles by reading mathematics, if you are a physicist, you would probably not succeed, because modern mathematics is extremely difficult to read, and I believe there exist only two kinds of modern mathematics books; one which you cannot read beyond the first page and one which you cannot read beyond the first sentence."

—C. N. Yang, *Lectures on Frontiers in Physics*, Seoul, Korea, 1980.

LARGE NUMBERS AND UNPROVABLE THEOREMS

JOEL SPENCER
Department of Mathematics, State University of New York, Stony Brook, NY 11794

"Yes, please," said Milo. "Can you show me the biggest number there is?"

"I'd be delighted," [the Mathemagician] replied, opening one of the closet doors. "We keep it right here. It took four miners just to dig it out."

Inside was the biggest

3

Milo had ever seen. It was fully twice as high as the Mathemagician.

—*The Phantom Tollbooth*
Norton Juster

1. Large Numbers. "Describe, on a 3 × 5 card, as large a positive integer as you can."

Many mathematicians have at some time played the game above, either solitaire or in competition. My solutions in the second, sixth, and twelfth grades, respectively, are shown in Figs. 1, 2, 3.

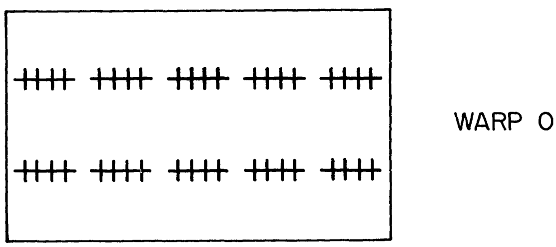


FIG. 1.

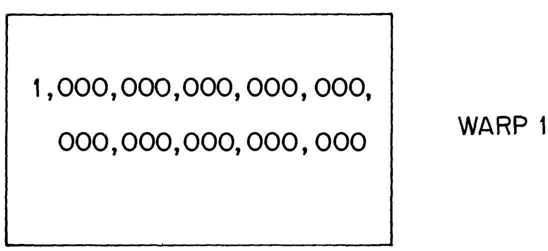


FIG. 2.

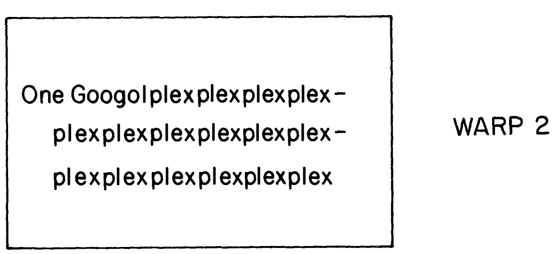


FIG. 3.

See this MONTHLY, 90 (1983)365, for the author's biography.

The last needs a word of explanation. Since googol is 10^{100} and googolplex is 10^{googol} let us define Nplex as 10^N. Actually, by twelfth grade I could write "One googolplexplexplex...with a googol plexes" and even some more elaborate variants. These were at best WARP 2.2. The next level is shown in Fig. 4.

Let $f_1(x) = 2x$
and $f_{n+1}(x) = f_n^{(x)}(1)$
$f_9(9)$

WARP 3

FIG. 4.

Here $f^{(x)}$ represents the xth iterate of f. Iterated doubling is exponentiation, $f_2(x) = 2^x$. Iterated exponentiation is the tower function,

$$f_3(x) = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}} \quad \text{with } x \quad 2\text{'s}.$$

My WARP 2 solution is approximately $f_3(21)$, one for each plex and five to get to a googol. There is no word for f_4. $f_4(4) = f_3(f_3(f_3(f_3(1)))) = f_3(f_3(4)) = f_3(65536)$ is already WARP 2.1.

Three ideas help us create large numbers. First, we concentrate on constructing rapidly growing functions. The numbers will then be the value of the function $f(x)$ for some reasonably small x. Second, we use iteration to build a larger function from a given one. Third, we introduce diagonalization. Having defined the functions f_n above, we define a diagonal function, called f_ω, by

$$f_\omega(n) = f_n(n).$$

This is called the Ackermann function. (There are several similar formulations.) The Ackermann function does occasionally appear in "real" mathematics. For example, van der Waerden proved in 1927 that to all n there exists $W(n)$ such that if the integers from 1 to $W(n)$ are divided into two classes, then there exists an arithmetic progression of length n in one of the classes. His proof gave a $W(n)$ roughly equal to $f_\omega(n)$. (It is possible that far smaller $W(n)$, even of exponential order, will suffice and this remains an open problem.)

Once $f_\omega(n)$ is defined, there is no reason to stop. We define a new function, let's call it $f_{\omega+1}$, by $f_{\omega+1}(n) = f_\omega^{(n)}(1)$. Having defined $f_{\omega+1}$, we may define $f_{\omega+2}, f_{\omega+3}, \ldots$. When faced with ellipses we resort to diagonalization. We define a new function, called $f_{2\omega}$, by $f_{2\omega}(n) = f_{\omega+n}(n)$. (See Fig. 5.)

$f_{2\omega}(9)$

WARP 3.2

FIG. 5.

We are defining here a hierarchy of functions in which each function has an immediate successor and where limit functions are defined by diagonalization of an appropriate subsequence. The usual representation for ordinal numbers provides a perfect framework in which to do this. The ordinals $\alpha < \omega^\omega$ have a simple representation. Each such α may be uniquely written

$$\alpha = a_1\omega^{s_1} + a_2\omega^{s_2} + \cdots + a_r\omega^{s_r} \qquad (\omega > s_1 > s_2 > \cdots > s_r \geqslant 0)$$

where the a_i are positive integers. (We write $a\omega^s$ instead of the more customary $\omega^s a$ for convenience of expression.) The limit ordinals are those α with $s_r > 0$. For these we define a specific "natural" sequence $\alpha(n)$ of ordinals approaching ω^ω*by*

$$\alpha(n) = a_1\omega^{s_1} + \cdots + a_{r-1}\omega^{s_{r-1}} + (a_r - 1)\omega^{s_r} + n\omega^{s_r - 1}.$$

For example, if $\alpha = 2\omega^4 + 3\omega^3$, then $\alpha(n) = 2\omega^4 + 2\omega^3 + n\omega^2$. We define the natural sequence approaching ω^ω by

$$\omega^\omega(n) = \omega^n.$$

Now we define $f_\alpha(n)$ for each $\alpha \leqslant \omega^\omega$ using transfinite induction by

$$f_{\alpha+1}(n) = f_\alpha^{(n)}(1), \tag{$+$}$$

$$f_\alpha(n) = f_{\alpha(n)}(n), \tag{$++$}$$

where α is a limit ordinal and the initial value $f_1(n) = 2n$. (See Fig. 6.)

$f_{\omega^\omega}(9)$ WARP 3.5

FIG. 6.

Let us emphasize that though we are using the language of infinite ordinals the functions f_α are recursive functions and the values $f_\alpha(t)$ are well-defined integers. The infinite ordinals are, in one sense, merely finite sequences of positive integers being manipulated in particular ways. A recursive program for computing $f_\alpha^{(t)}(n)$ could take the following form.

```
FUNCTION F(α,N,T)
BEGIN
IF T > 1,
          SET X = F(α,N,T − 1)
          RETURN F(α,X,1)
IF T = 1 AND α = 1
          RETURN 2*N
IF T = 1 AND LIMITORDINAL (α)
          RETURN F(α(N),N,1)
IF T = 1 AND NOT LIMITORDINAL (α)
          RETURN F(α − 1,1,N)
END
```

The representation of α, the predicate LIMITORDINAL (α), and the functions $\alpha - 1$ and $\alpha(N)$ need to be defined explicitly, though we do not do so here.

We continue the ordinals a half-WARP further. Set

$$\omega_1 = \omega,\ \omega_2 = \omega^\omega, \ldots, \omega_{s+1} = \omega^{\omega_s}, \ldots$$

and set ε_0 equal the limit of the ω_s. (We emphasize that ω_1 is *not* the first uncountable ordinal. All ordinals in this paper are countable.) Each ordinal $\alpha < \omega_{s+1}$ is uniquely represented as

$$\alpha = a_1\omega^{\beta_1} + \cdots + a_r\omega^{\beta_r} \qquad (\omega_s > \beta_1 > \beta_2 > \cdots > \beta_r \geqslant 0)$$

with the a_i positive integers. A "typical" ordinal is

$$7\omega^{\omega^{2\omega+1}} + 14\omega^{3\omega^{\omega+8}+5\omega^\omega}$$

Now for limits. We say $n\omega^{\beta}$ is the natural sequence approaching $\omega^{\beta+1}$. If β itself is a limit ordinal, then its limit sequence $\beta(n)$ has already been defined and we call $\omega^{\beta(n)}$ the natural sequence approaching ω^{β}. For sums we keep all but the smallest term fixed and take a limit sequence approaching that smallest term. Thus

$$7\omega^{\omega^{2\omega+1}} + 13\omega^{3\omega^{\omega+8}+5\omega^{\omega}} + \omega^{3\omega^{\omega+8}+4\omega^{\omega}+\omega^{n}}$$

is the natural sequence for the ordinal above. Finally, ε_0 has the natural sequence $\varepsilon_0(n) = \omega_n$. Now the hierarchy f_α defined by $(+)$, $(++)$ may be extended to all $\alpha < \varepsilon_0 + \omega$. We have a big number. (See Fig. 7.)

$$\boxed{f_{\epsilon_0+9}(9)}$$

WARP 4

FIG. 7.

This should win the game against any nonlogician!

2. The Connection. Let PA stand for Peano Arithmetic, that first order theory of numbers which includes the basically finitistic methods of number theory. The surprising truth is that WARP 4 lies beyond the scope of PA. The sense in which we use this was shown by G. Kreisel [**4**] in 1952.

A statement $P(x_1,\ldots,x_r)$ is called provably recursive if there is an algorithm for deciding if $P(x_1,\ldots,x_r)$ is true and a proof, in PA, that the algorithm always terminates. Thus $P(x, y, z, t): x^t + y^t = z^t$ is provably recursive (simply make the calculation) but

$$P(t)\colon (Ex)(Ey)(Ez)x^t + y^t = z^t$$

is not known to be provably recursive.

We say a function f dominates a function g if there exists n such that $f(x) > g(x)$ for all $x > n$.

Let $P(x, y)$ be a provably recursive statement in PA and suppose $(x)(Ey)P(x, y)$ is provable in PA. Set $f_P(x)$ equal the least y such that $P(x, y)$ is true. Then, Kreisel showed, the function f_P is dominated by f_α for some $\alpha < \varepsilon_0$. As f_{ε_0} dominates all previous f_α we draw the following conclusion.

> Let $(x)(Ey)P(x, y)$ be a statement of PA which is true for the natural numbers and let $f_P(x)$ be the least y for which $P(x, y)$ is true. Suppose $P(x, y)$ is provably recursive. If f_P dominates f_{ε_0}, then the statement $(x)(Ey)P(x, y)$ is unprovable in PA.

3. An Unprovable Theorem.* The epochal work of Kurt Gödel gave the existence of statements in PA which are true for the natural numbers but unprovable in PA. The statements constructed by Gödel suffered the defect of being unnatural and for the past half century a somewhat raggedy debate ensued concerning whether or not Gödel's result applied to statements of real mathematical interest. In 1977 Jeff Paris and Leo Harrington [**2**] gave the first natural example of a statement that was true for the integers and unprovable in PA. (The term "natural" is here a matter of subjective opinion.) Their statement comes from Ramsey Theory, a subdisci-

*The term "unprovable theorem" is abhorred by logicians. Theorems have proofs by definition. For our informal discussion, however, it seems appropriate to the subject matter to use this delightful oxymoron.

pline of Combinatorial Analysis, and to give it one needs a moment's introduction to that subject. (A detailed treatment is given in [**1**].)

By "an r-coloring of the k-sets of S" we mean a function χ with domain the family of k-element subsets of S and range $[r]$. (Notation: $[a, b] = \{a, a+1, \dots, b\}, [r] = [1, r], [a, b) = [a, b-1]$.) Given such a coloring χ a set $B \subset S$ is called monochromatic if all of the k-element subsets of B have the same color.

We may state Ramsey's Theorem in either a finite or an infinite form.

RAMSEY'S THEOREM (*Infinite Form*). *For all k, r given any r-coloring of the k-sets of N, there exists a monochromatic infinite set B.*

RAMSEY'S THEOREM (*Finite Form*). *For all k, r, t there exists n so that given any r-coloring of the k-sets of $[n]$, there exists a monochromatic t-set B.*

From the infinite form of Ramsey's Theorem we deduce the finite form as follows. Suppose the finite form false and fix k, r, t so that for all n there exists an r-coloring of the k-sets of $[n]$ with no monochromatic t-set B. Any coloring for a larger n also works for a smaller one. Hence, for any given n, there is some coloring which can be extended to arbitrarily large n. Construct colorings for successively larger values of n in turn, each of which extends to arbitrarily large n. The union is an r-coloring of the k-sets of N with no monochromatic t-set B. Thus the infinite form of Ramsey's Theorem would be false. The reasoning above, often called a Compactness Argument, can be applied in many situations to reduce an "infinite form" to a "finite form," see, e.g., [**1**].

Define a set S of positive integers to be large if $|S| > \min(S)$. For example, $\{3,4,7,9\}$ is large but $\{4,63,1281,4504655\}$ is not. The statement of Paris and Harrington (in one version) requires a seemingly minor modification of Ramsey's Theorem.

(PH) *For all k, r there exists n so that given any r-coloring of the k-sets of $[k+1, n]$ there exists a large monochromatic $B \subset [k+1, n]$.* (The exclusion of $1, \dots, k$ is purely technical, avoiding such trivial large sets as $\{2,3,4\}$.)

If we allow infinitistic techniques, (PH) is relatively simple to prove. Suppose (PH) is false for a particular k, r. By the Compactness Argument there would exist an r-coloring χ of the k-sets of $[k+1, \infty)$ with no large monochromatic finite B. However, given any such χ the infinite form of Ramsey's Theorem guarantees the existence of an infinite monochromatic set C. The first $\min(C)+1$ elements of C then give a large monochromatic finite B.

We have deduced both Ramsey's Theorem (finite form) and (PH) from Ramsey's Theorem (infinite form). Neither of these arguments is formalizable in PA since neither Ramsey's Theorem (infinite form) nor the Compactness Argument can even be stated in PA. This, by itself, does not show that Ramsey's Theorem (finite form) or (PH) are unprovable in PA, only that we have not proven them. In fact, Ramsey's Theorem (finite form) can be proven in PA (though we do not prove it here) but (PH) cannot.

Paris and Harrington, in their original work, showed by model-theoretic arguments that (PH) was unprovable in PA. Robert Solovay, hearing of their result but not of their proof, discovered a more combinatorial argument. Let $\mathrm{PH}(k, r)$ be the least n such that for every r-coloring of the k-element subsets of $[k+1, n]$ there exists a monochromatic large B. Solovay showed that PH grows too fast for PA.

A full discussion of Solovay's argument is somewhat beyond the bounds of this expository discussion (though not by too much, see [**1**] or the original [**3**]), but we can quite easily demonstrate that $\mathrm{PH}(2, r)$ grows quite rapidly. (A similar exposition was given by Smoryński [**5**].) To find a lower bound for $\mathrm{PH}(2, r)$ we give explicit r-colorings of the 2-sets of $[3, n]$.

Split $[3, \infty)$ into consecutive intervals of the form $[x, 2x)$—i.e., $[3,6), [6,12)\ [12,24), [24,48), \dots$. We give the pair $\{i, j\}$ color 1 if i and j lie in a common interval. If all pairs in a set

$A = \{a_1, \ldots, a_s\}$ have color 1, then $A \subset [x, 2x)$, so $|A| \leqslant x$ and $\min(A) \geqslant x$, hence A is not large. Set $g_1(x) = 2x$. Now we define $g_2(x) = g_1^{(x)}(x) = x2^x$ and split $[3, \infty)$ into consecutive intervals of the form $[x, g_2(x))$—that is,

$$[3,24),\ [24, 24 \cdot 2^{24}),\ [24 \cdot 2^{24}, 24 \cdot 2^{24} \cdot 2^{24 \cdot 2^{24}}), \ldots.$$

We give a pair $\{i, j\}$ color 2 if i and j lie in a common interval and the pair does not have color 1. If all pairs in a set $A = \{a_1, \ldots, a_s\}$ have color 2, then $A \subset [x, g_2(x))$, which is split into x subintervals. Each element of A lies in a separate subinterval (since no pair has color 1) so $|A| \leqslant x$ and A is not large. On $[3, g_2(3))$ all 2-sets have either color 1 or 2 and there are no monochromatic large sets. Thus

$$\mathrm{PH}(2,2) \geqslant g_2(3) = 24.$$

We continue in this manner, defining $g_{s+1}(x) = g_s^{(x)}(x)$, partitioning $[3, \infty)$ into consecutive intervals of the form $[x, g_{s+1}(x))$, and giving a pair $\{i, j\}$ color $s + 1$ if i and j lie in a common interval and the pair has not been given a smaller color. Then, quite explicitly, we have shown

$$\mathrm{PH}(2,3) \geqslant g_3(3) = 24 \cdot 2^{24} \cdot 2^{24 \cdot 2^{24}}$$

and, in general, $\mathrm{PH}(2, r) \geqslant g_r(3)$. The function $g_r(3)$ has order roughly $f_\omega(r)$.

The colorings of k-sets are equally explicit but require a greater technical effort. Solovay showed that $\mathrm{PH}(3, r)$ is bounded from below by $f_\alpha(r)$ where $\alpha = \omega^\omega$, $\mathrm{PH}(4, r)$ by $f_\alpha(r)$ where $\alpha = \omega^{\omega^\omega}$, etc., and that $\mathrm{PH}(r, r)$ was bounded from below by $f_{\varepsilon_0}(r)$. (Though we do not require it here, Ketonen found upper bounds on these functions of roughly the same order.)

Let $P(k, n)$ be the statement "Given any k-coloring of the k-sets of $[k + 1, n]$ there exists a large monochromatic B." $P(k, n)$ is surely provably recursive as one may check all k-colorings of the k-sets of $[k + 1, n]$. Applying Kreisel's fundamental result the statement

$$(k)(En)P(k, n)$$

is unprovable in PA.

4. Reflections. WARP 4 takes us to the tradeoff between largeness and definiteness. We have described an algorithm for computing $f_\alpha(t)$—but how do we know that the algorithm will work (i.e., terminate)? One way is by transfinite induction, the determination of $f_\alpha(t)$ requires t calls of the algorithm to calculate $f_{\alpha-1}$ or, if α is a limit ordinal, the algorithm for $f_{\alpha(t)}$. In either case these are smaller ordinals, by induction the algorithm works for them, hence the f_α algorithm works. However, transfinite induction is a basically infinitistic tool and we can ask for a proof in PA that the algorithm for f_α will work. Here there is a very nice result. For $\alpha < \varepsilon_0$ there is such a proof in PA. However, for $\alpha = \varepsilon_0$ there is no proof in PA that the algorithm will always work. (This gives another statement which is true but unprovable, but one that would hardly be termed natural.)

We would agree that the number described in Fig. 8 is not legitimate as it stands. It gives, in fact, Berry's paradox, one of the classic Russell-type paradoxes. Since this number has been described in less than 50 words, it must be greater than itself. The problem lies in the notion of describable. Let us say that a number m is describable (modulo PA) in length n if there is a statement $A(x)$ in PA such that

(i) $(E_1 x)A(x)$ has a proof in PA of length at most n. (E_1 = "there exists a unique.")
(ii) There is an algorithm for deciding $A(x)$ and a proof in PA of length at most n that the algorithm always terminates.
(iii) $A(m)$.

We construct a legitimate alternative (see Fig. 9). This number cannot be described by a book the size of the known universe, with electrons for characters, in the language of PA. The function g lies beyond PA—but just barely. It is of order f_{ε_0} and the above card is still WARP 4.

To go beyond WARP 4 we strengthen PA. Let GAM be a formalization of Generally Accepted Mathematics. (See Fig. 10.)

The largest number describable in at most fifty words, plus one.

WARPED

FIG. 8.

Let $g(n)$ equal the largest m describable (modulo PA) in length $\leq n$.

$g(10^{100}) + 1$

FIG. 9.

Let $g_{GAM}(n)$ equal the largest m describable (modulo GAM) in length $\leq n$.

$g_{GAM}(10^{100}) + 1$

WARP 5 (?)

FIG. 10.

Travel beyond WARP 4 now depends on what one allows in GAM. There is always the danger that if too much is allowed, the system will become inconsistent and the 3×5 card will no longer define an integer. The game of describing the largest integer, when played by experts, lapses into hopeless argument over legitimacy.

References

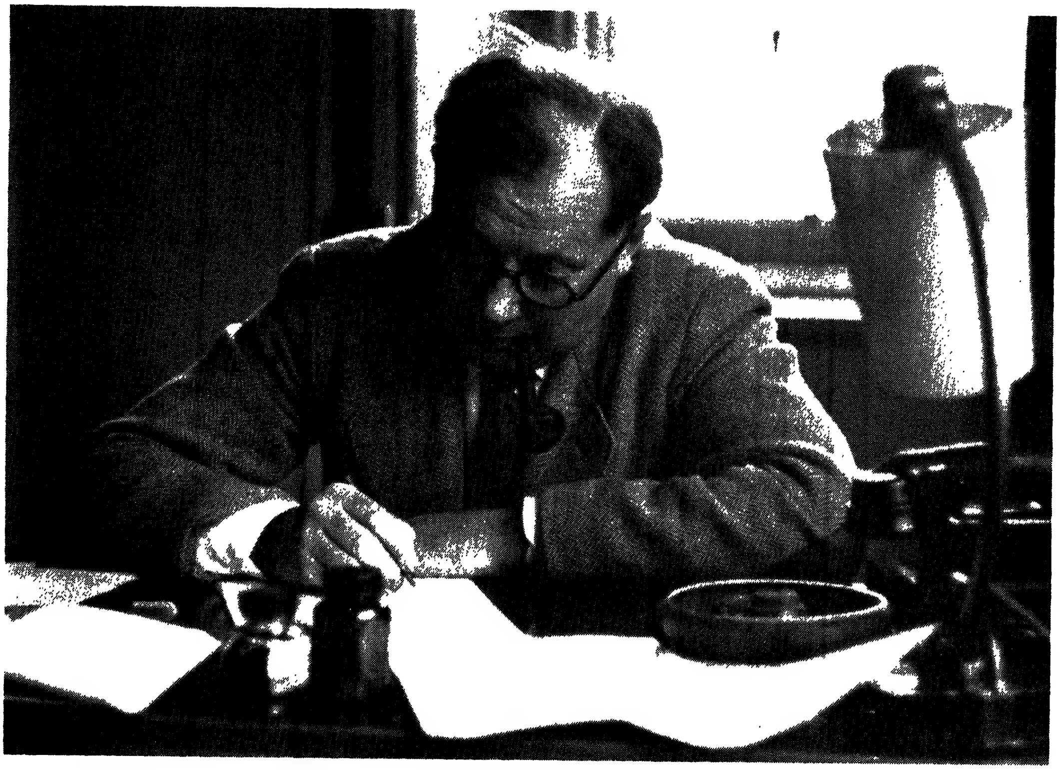
1. R. L. Graham, B. L. Rothschild, and J. H. Spencer, Ramsey Theory, Wiley, New York, 1980.
2. J. Paris and L. Harrington, A Mathematical Incompleteness in Peano Arithmetic, in Handbook of Mathematical Logic (J. Barwise, Editor), North-Holland, 1977, 1133–1142.
3. J. Ketonen and R. Solovay, Rapidly growing Ramsey functions, Ann. of Math., 113 (1981)267–314.
4. G. Kreisel, On the interpretation of nonfinitistic proofs, II, J. Symbolic Logic, 17 (1952)43–58.
5. C. Smoryński, Some rapidly growing functions, Math. Intelligencer, 2 (1980)149–154.

MISCELLANEA

116.

The condensation of metaphor involves no denial of logic: it is simply an extension of the implications of grammar, the development of a notation which, being less cumbersome, enables us to think more easily. It may be compared to the invention of a new notation, say that of Leibniz or Hamilton, in mathematics: the new is defined in terms of the old, it is a shorthand which must be learned by patient effort, but, once learnt, it makes possible the solution of problems which were too complicated to attack before. The human head can only carry a certain amount of notation at any one moment and poetry takes up less space than prose.

–Michael Roberts, *The Faber Book of Modern Verse*, London, Faber and Faber, 1937, p. 20.



His uncle is better known to most educated people, but for topologists he is the important one. (See p. 705.)

squared error as a measure of quality. It is well known and easy to verify that

$$\mathrm{MSE}\left(\frac{n}{n+1}S^2(\bar{X})\right) \leqslant \mathrm{MSE}\left(anS^2(\bar{X})\right)$$

and

$$\mathrm{MSE}\left(\frac{n}{n+2}S^2(\mu)\right) \leqslant \mathrm{MSE}\left(anS^2(\mu)\right)$$

for every $a > 0$. Thus if mean squared error is our criterion, we should use

$$\frac{n}{n+2}S^2(\mu) = \frac{1}{n+2}\sum(X_i-\mu)^2$$

when μ is known (an admissible estimator) and

$$\frac{n}{n+1}S^2(\bar{X}) = \frac{1}{n+1}\sum(X_i-\bar{X})^2$$

when μ is unknown (an inadmissible estimator (cf. Stein (1964)). Moreover, our intuition is then confirmed as

$$\mathrm{MSE}\left(\frac{n}{n+2}S^2(\mu)\right) = \frac{2}{n+2}\sigma^4 < \frac{2}{n+1}\sigma^4 = \mathrm{MSE}\left(\frac{n}{n+1}S^2(\bar{X})\right).$$

Thus the statistic using the known mean μ is better, in the sense of mean squared error, than the one ignoring this information. It should be noted, however, that these two estimators of σ^2 are almost never considered in basic textbooks on statistics.

On the other hand, one might not be so much interested in σ^2 as in σ. Thus we might compare the mean squared errors of $S(\bar{X})$ and $S(\mu)$ as estimators of σ. The distributions of these statistics are not so tractable as those of $S^2(\bar{X})$ and $S^2(\mu)$. However, we calculated their mean squared errors for sample sizes of $1, 2, \ldots, 30, 40, 50, \ldots, 100$, and in each of these instances the mean squared error of $S(\mu)$ was less than that of $S(\bar{X})$. Thus it seems that for the estimators $S(\bar{X})$ and $S(\mu)$ of the standard deviation σ our intuition holds while, for the estimators $S^2(\bar{X})$ and $S^2(\mu)$ of σ^2 it does not hold.

Finally, mean squared error is not the only possible criterion for measuring the quality of an estimator. Among other measures of the "closeness" of an estimator to its parameter is one proposed by Pitman. It can be shown that

$$P\left[\left(S^2(\bar{X})-\sigma^2\right) > \left(S^2(\mu)-\sigma^2\right)\right] \geqslant \frac{1}{2}$$

so that, in the Pitman sense, $S^2(\mu)$ is "closer" to σ^2 than $S^2(\bar{X})$ is. Here again our intuition is upheld.

References

1. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

2. M. G. Kendall, The Advanced Theory of Statistics, C. Griffen & Co., 3rd ed., vol. 2, 1946.

3. D. A. Pierce, The asymptotic effect of substituting estimators for parameters in certain types of statistics, Ann. Statist., 10 (1982) 475–478.

4. R. H. Randles, On the asymptotic normality of statistics with estimated parameters, Ann. Statist., 10 (1982) 462–474.

5. C. Stein, Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Ann. Inst. Statist. Math., 16 (1964) 155–160.

ANSWER TO PHOTO ON PAGE 676

J.H.C. Whitehead.

FUNCTIONS WHICH PARAMETRIZE MEANS

MICHAEL E. MAYS
Department of Mathematics, West Virginia University, Morgantown, WV 26506

1. Introduction. This paper is about certain functions of two variables, means, and several natural ways that have arisen to associate them with functions of a single variable.

We define a mean $m(a, b)$ to be a function of two positive variables satisfying the following properties:

$$\text{(Intermediacy)} \quad \min(a, b) \leqslant m(a, b) \leqslant \max(a, b) \tag{1}$$

$$\text{(Symmetry)} \qquad m(a, b) = m(b, a). \tag{2}$$

Several means well-enough known to have names are listed below:

Arithmetic	$A(a, b) = (a + b)/2$
Geometric	$G(a, b) = \sqrt{ab}$
Harmonic	$H(a, b) = 2ab/(a + b)$
Logarithmic	$L(a, b) = (b - a)/(\ln b - \ln a)$
Root Mean Square	$RMS(a, b) = [(a^2 + b^2)/2]^{1/2}$
Contraharmonic	$C(a, b) = (a^2 + b^2)/(a + b)$
Heronian	$He(a, b) = (a + \sqrt{ab} + b)/3$
Identric	$I(a, b) = \left(\dfrac{a^a}{b^b}\right)^{1/(a-b)} \Big/ e$
Maximum	$\mathrm{Max}(a, b)$
Minimum	$\mathrm{Min}(a, b)$.

The logarithmic mean has been studied by Lin [**12**] and Carlson [**3**]. It was generalized by Stolarsky [**10**, **11**], and studied by him and Leach and Sholander [**7**]. A special case of the generalization is the identric mean—it also arises as the value of a certain limit on the 1979 Putnam Examination.

There is a technique called compounding which builds new means from old. Carlson [**2**] develops this iterative procedure whereby (possibly nonsymmetric) functions of two variables are combined to yield a new mean and generalizes Gauss's algorithm for the arithmetic-geometric mean and Borchardt's algorithm for the calculation of the inverse sine. Lehmer [**8**] discusses the compounding of means from certain families by associating the families with series expansions. Gould and Mays [**5**] derive recurrence relations and explicit formulas for the coefficients in the series expansions of the families Lehmer studied and others.

All of the named means satisfy a third property as well.

$$\text{(Homogeneity)} \quad m(a, b) = am(1, b/a) \tag{3}$$

for all a and b. We will speak of a homogeneous mean if we wish (3) to hold. One goal of this paper is to determine conditions on a function so that the mean associated with that function is homogeneous. Another aim is to find functions that are associated with the means mentioned at the beginning. The techniques we consider for associating means with functions of a single variable include a graphical technique of Moskovitz and an application of the Mean Value Theorem for integrals which is related to the approach of Stolarsky [**10**, **11**] and the classical work of Hardy, Littlewood, and Pólya [**6**].

2. The alignment chart of Moskovitz. Let f be any function from $(0, \infty)$ into $(0, \infty)$, and define

---

Michael E. Mays: I received my Ph.D. from Penn State in 1977 as a student of Raymond Ayoub. Since then I have been at West Virginia University. Besides collecting means, my research interests include estimating the rate of growth of counting functions associated with properties of finite groups.

$M_f(a, b)$ to be the x intercept of the line connecting $(a, f(a))$ and $(b, -f(b))$. It is easy to see that intermediacy and symmetry both hold for M_f. To find the mean associated with a given f, calculate the slope of the line through $(a, f(a))$, $(M_f(a, b), 0)$, and $(b, -f(b))$ in two ways to obtain (see Fig. 1)

$$\frac{f(a)}{a - M_f(a, b)} = \frac{f(b)}{M_f(a, b) - b}. \tag{4}$$

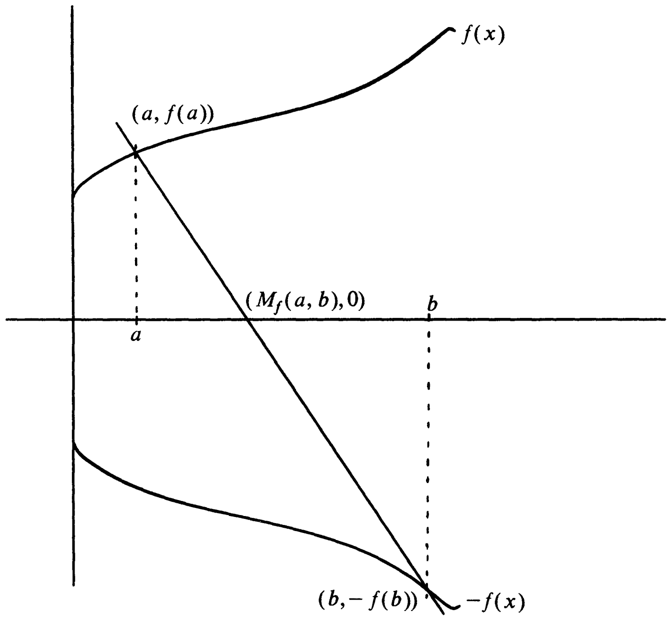


FIG. 1.

This can be written as

$$M_f(a, b) = \frac{af(b) + bf(a)}{f(a) + f(b)}. \tag{5}$$

Different functions can determine the same mean.

THEOREM 1. *$M_f = M_g$ if and only if $g = kf$ for some $k > 0$.*

Proof. If $g = kf$, k cancels in the right-hand side of (5) to yield $M_f = M_g$. If $g \neq kf$ for any k, pick a, b, and k so that $g(a) = kf(a)$ but $g(b) \neq kf(b)$. Then if $M_f(a, b) = M_g(a, b)$,

$$\frac{af(b) + bf(a)}{f(a) + f(b)} = \frac{ag(b) + bkf(a)}{kf(a) + g(b)}. \tag{6}$$

Cross-multiplying and regrouping yields

$$af(a)[kf(b) - g(b)] = bf(a)[kf(b) - g(b)],$$

and since $kf(b) - g(b) \neq 0$ and $f(a) \neq 0$, we must have $a = b$, a contradiction.

COROLLARY. *When associating M_f with a given function f, we may assume without loss of generality that $f(1) = 1$.*

Any function f determines a mean M_f by formula (5), but not every mean is determined by a

function. There are obvious difficulties with Max and Min, and less obvious difficulties with some other homogeneous means.

THEOREM 2. *Let $m(a, b)$ be a homogeneous mean, and define $f(x)$ by*

$$f(x)=\begin{cases}\dfrac{x-m(1,x)}{m(1,x)-1} & x\neq 1,\\ 1 & x=1.\end{cases} \tag{7}$$

If $f(x)$ is multiplicative, then $m = M_f$. If $f(x)$ is not multiplicative, then $m \neq M_f$ for any f.

Proof. If $m = M_f$, f is determined by (5) with $a = 1$, $b = x$, and $f(1) = 1$ to be (7). Since m is homogeneous,

$$m(x, xy) = xm(1, y),$$

so

$$\frac{xf(xy) + xyf(x)}{f(x) + f(xy)} = x\frac{f(y) + y}{1 + f(y)},$$

and cross-multiplying again gives an equation which simplifies to

$$f(xy) = f(x)f(y).$$

This says that the only possible functions yielding homogeneous M_f are constant multiples of powers of x. It is to be emphasized that as long as $m(1, x) \neq 1$ the function $f(x)$ given by (7) may be used to calculate correctly the mean of 1 and x. It is only when neither argument of the mean is 1 that the trouble arises.

For example, (7) associates with the Heronian mean the function

$$f(x) = \frac{x - (x + \sqrt{x} + 1)/3}{(x + \sqrt{x} + 1)/3 - 1} = \frac{1 + 2\sqrt{x}}{2 + \sqrt{x}},$$

which is not multiplicative.

$$He(3,7) = (3 + \sqrt{21} + 7)/3 = 4.861, \text{ but}$$

$$\frac{3f(7) + 7f(3)}{f(3) + f(7)} = 4.876.$$

Of the means considered by Moskovitz, only the following have multiplicative functions associated with them by (7):

M_f	f
A	1
G	$\sqrt{x}$
H	x
C	$1/x$.

The technique of Moskovitz provides a set of homogeneous means parametrized by a single variable, given by

$$M_s(a, b) = M_{x^s}(a, b) = \frac{ab^s + ba^s}{a^s + b^s}. \tag{8}$$

Either a geometric or an analytic argument gives that, for fixed a and b, M_s is a decreasing function of s; $\lim_{s\to\infty} M_s = \text{Min}$ and $\lim_{s\to-\infty} M_s = \text{Max}$. The property of being monotone in the parameter is shared by several other sets of means that have occurred in the literature, among them the power means

$$P_s(a, b) = [(a^s + b^s)/2]^{1/s}, \tag{9}$$

a mean of Gini [4], also studied by Beckenbach [1] and Lehmer [8]

$$G_s(a, b) = \frac{a^s + b^s}{a^{s-1} + b^{s-1}} \tag{10}$$

or Stolarsky's generalized logarithmic mean

$$U_s(a, b) = \left\{ \frac{a^s - b^s}{s(a - b)} \right\}^{1/(s-1)}. \tag{11}$$

That $M_s(a, b)$ and $G_s(a, b)$ have similar properties is explained by the identity

$$M_s(a, b) = G_{1-s}(a, b).$$

Two parameter families also exist. Several of Gini's means of several variables reduce in the case of two variables to

$$G(r, s; a, b) = \left\{ \frac{a^s + b^s}{a^r + b^r} \right\}^{1/(s-r)}, \tag{12}$$

which generalizes (10), and Leach and Sholander worked with

$$E(r, s; a, b) = \left\{ \frac{r(a^s - b^s)}{s(a^r - b^r)} \right\}^{1/(s-r)}, \tag{13}$$

which generalizes (11). (9) is a special case of both (12) and (13). The two parameter means are monotone in each parameter if a and b and the other parameter are fixed.

3. Means and Mean Value Theorems. Formula (11) arises as a special case of a general method for associating a mean with a function in [**10**], just as (8) arises from the technique in [**9**]. Let g be (strictly) monotone, differentiable, and convex on $(0, \infty)$. Then the Mean Value Theorem for derivatives guarantees that for each pair of numbers a and b there exists a unique c in (a, b) satisfying

$$g'(c) = (g(b) - g(a))/(b - a). \tag{14}$$

Write $c = U_g(a, b)$. The c chosen is clearly between a and b, and symmetry is also obvious (Fig. 2).

The Mean Value Theorem for integrals generates mean values, too. We need only assume that f is (strictly) monotone and continuous on $(0, \infty)$ to be sure that a unique c in (a, b) exists satisfying

$$f(c)(b - a) = \int_a^b f(x)\, dx. \tag{15}$$

We define the mean V_f by

$$V_f(a, b) = f^{-1}\left[\int_a^b f(x)\, dx/(b - a)\right]. \tag{16}$$

The next theorem shows that the sets of means

$$\{U_g | g \text{ is monotone, differentiable, and convex}\}$$

and

$$\{V_f | f \text{ is monotone and continuous}\}$$

are the same, but we will use the V_f form exclusively thereafter because a broader class of functions are allowed to generate means (at least superficially), there is a more elegant characterization of those functions which generate the same mean, the integral has a clearer link to the work of Hardy, Littlewood, and Pólya, and all the "nice" means are associated with powers of x (Fig. 3).

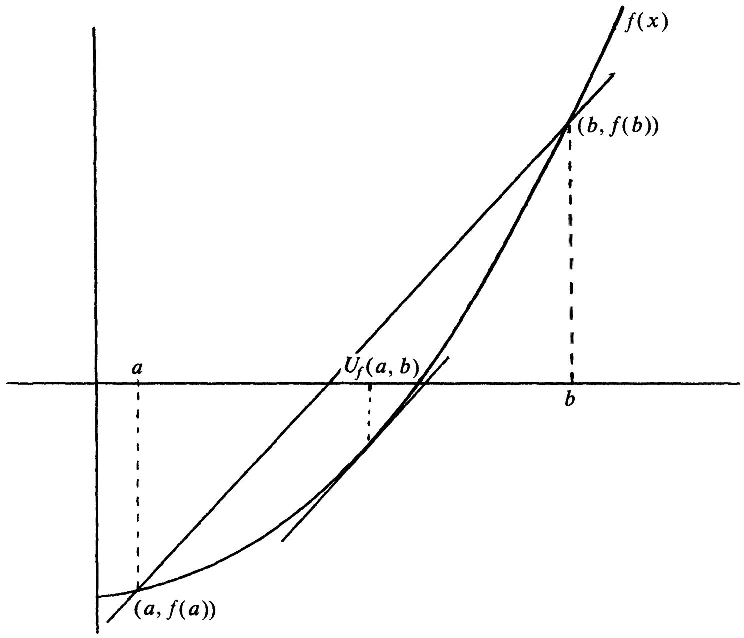


FIG. 2.

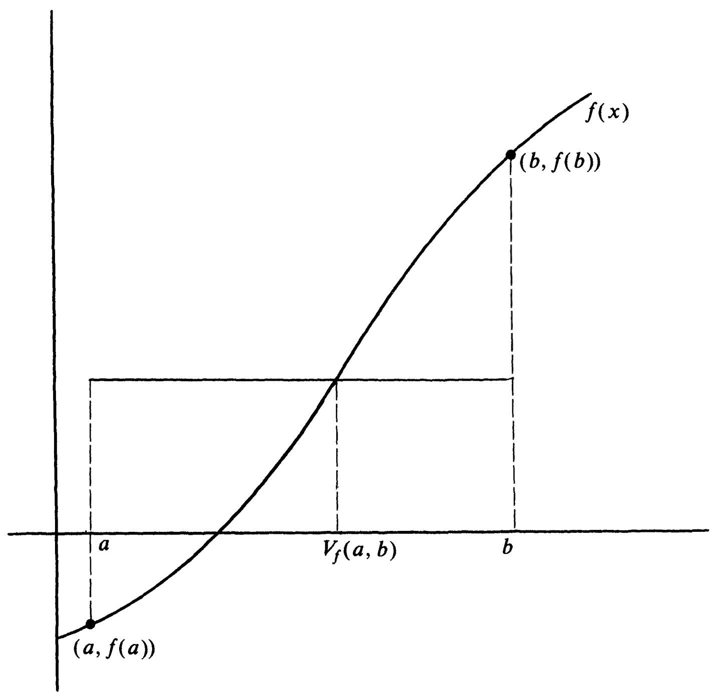


FIG. 3.

THEOREM 3. *If f is monotone and continuous, then*

$$V_f = U_{\int f(x)\,dx}.$$

If g is monotone, differentiable, and convex, then

$$U_g = V_{D_x g}.$$

The proof is an application of properties of real functions and the Fundamental Theorem of Calculus.

Hardy, Littlewood, and Pólya associate a mean of several numbers with a continuous, monotone function. For the n nonnegative numbers $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and weights $q_1, q_2, \ldots, q_n$ which sum to 1, $\mathfrak{M}_f(a_1, a_2, \ldots, a_n)$ is defined by

$$\mathfrak{M}_f(a_1, a_2, \ldots, a_n) = f^{-1}\left[\sum_{i=1}^{n} q_i f(a_i)\right].$$

Take $a_1 = a$, $a_2 = a + (b - a)/n, \ldots$, and $a_n = a + (b - a)(n - 1)/n$, and $q_i = 1/n$ for each i to get

$$\sum_{i=1}^{n} q_i f(a_i)$$

a Riemann sum for $\int_a^b f(x)\,dx/(b - a)$. It is an upper sum if f is decreasing and a lower sum if f is increasing. Fancier choices for the q_i yield a Stieltjes integral in the limit. In a sense, we have constructed a mean of two variables as a generalization of a mean of n variables. Properties of $\mathfrak{M}_f$ developed in [6] yield properties of V_f. Theorem 16 in [6] implies that, for V_{x^s} abbreviated as V_s, V_s is an increasing function of s for fixed a and b. We list some means arising as a V_s for some s.

s	V_s
$\to \infty$	Max
1	A
$\to 0$	I
$-1/2$	$(A + G)/2$
-1	L
-2	G
-3	$(HG^2)^{1/3}$
$\to -\infty$	Min

Conspicuous in its absence from this list is H, although the mean arising for $s = -3$ is enough to exhibit that $H \leqslant G$. No matter what f is taken, H never arises. To see why, carry the parallel between this construction and that of Hardy, Littlewood, and Pólya a few theorems further.

Theorem 83 of [6] implies the following result:

THEOREM 4. *$V_f(a, b) = V_g(a, b)$ for all a and b if and only if there exist constants $c \neq 0$ and k such that $f = cg + k$.*

A consequence of this theorem is that we may assume f is increasing without loss of generality. In fact we may take $f(1) = 1$ and, if f is differentiable at $1, f'(1) = 1$ by an appropriate choice of c and k.

The analogue of Theorem 84 of [6] provides a criterion for homogeneity.

THEOREM 5. *Suppose f is continuous and monotone on $(0, \infty)$, and that V_f is homogeneous. Then $V_f = V_s$ for some real number s.*

THEOREM 6. *$H \neq V_f$ for any f.*

Proof. By Theorem 5, since H is homogeneous we need only look among the means V_s. If

$V_s(1,2) = H(1,2) = 4/3$, solving

$$\left[(2^{s+1} - 1)/(s+1)\right]^{1/s} = 4/3$$

yields $s = -5.14$, but if $V_s(1,3) = H(1,3) = 3/2$,

$$\left[(3^{s+1} - 1)/(2s+2)\right]^{1/s} = 3/2$$

has solution $s = -5.36$.

References

1. E. F. Beckenbach, A class of mean value functions, this MONTHLY, 57 (1950) 1–6.
2. B. C. Carlson, Algorithms involving arithmetic and geometric means, this MONTHLY, 78 (1971) 496–505.
3. ______, The logarithmic mean, this MONTHLY, 79 (1972) 615–618.
4. C. Gini, Di una formula comprensiva delle medie, Metron, 13 (1938) 3–22.
5. H. W. Gould and M. E. Mays, Series expansions of means, to appear in J. Math. Anal. Appl.
6. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge, 1952.
7. E. B. Leach and M. C. Sholander, Extended mean values, this MONTHLY, 85 (1978) 84–90.
8. D. H. Lehmer, On the compounding of certain means, J. Math. Anal. Appl., 36 (1971) 183–200.
9. David Moskovitz, An alignment chart for various means, this MONTHLY, 40 (1933) 592–596.
10. K. B. Stolarsky, Generalizations of the logarithmic mean, Math. Mag., 48 (1975) 87 –92.
11. ______, The power and generalized logarithmic means, this MONTHLY, 87(1980) 545–548.
12. Tung-Po Lin, The power mean and the logarithmic mean, this MONTHLY, 81 (1974) 879–883.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

MONTHLY UNSOLVED PROBLEMS 1969–1983

RICHARD K. GUY

Department of Mathematics & Statistics, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4

Regular readers will know that in December of odd-numbered years we pool the information that correspondents send in concerning problems that have appeared in this section. References in brackets are to years and pages of this MONTHLY, earlier updating articles being [1971, 1113; 1973, 1120; 1975, 995; 1977, 807; 1979, 847 and 1981, 775]. In several of these there are statements of policy, including that of not normally publishing solutions. A parenthesis with a year after a name, such as Guy (1981), relates to the list of references at the end, as do (tbp) and (wrc), which respectively mean "to be published" and "written communication."

To accommodate a wider variety of unsolved problems, we have experimented with different formats, and sometimes, when there's a dearth of suitable problems, with no format at all!

There would not be enough space to pass on all the information if we restated all the problems in detail, so that many are given only a brief mention, without even defining the main ideas involved. This might betoken a regrettable lack of leisureliness. On the other hand, if we are encouraged to turn to our back issues, we are reminded of the wealth of good material that they contain.

Falconer (1983) has settled a problem of Klee [1969, 54; correction 1971, 1114] by showing that, except for certain unlikely possibilities, any curve with two equireciprocal points must have the same equireciprocal constant at each point. A point P is an equireciprocal point of a curve C if C is starshaped at P and if every chord XY of C through P satisfies $1/|XP| + 1/|PY| = \alpha$ for some (equireciprocal) constant α. Further, any twice differentiable curve with two equireciprocal points must be an ellipse. On the other hand, there are nonelliptical convex curves with two equireciprocal points. Hallstrom (1974) had obtained some partial results.

If $\phi(x)$ is Euler's totient function, then Carmichael's conjecture is that for no value of n does the equation $\phi(x) = n$ have a unique solution: see Klee [1969, 288]. Carmichael showed that such an n is greater than 10^{37} and Klee in 1944 improved this to 10^{400}. Now Masai and Valette (1982) have a lower bound of 10^{10000}. Valette (1981) has also published an expository paper.

The quest for gracefulness [1969, 1128] has abated somewhat, we are glad to say. All seekers are urged to read the paper of Huang, Kotzig and Rosa (1982). Ayel and Faveron (1982) have shown that helms (wheels with an additional edge at each rim-vertex) are graceful (Fig. 1). Golomb (1981) rides a number of hobby-horses, including polyominoes and graceful numbering, but there is no progress with the main Ringel-Kotzig conjecture.

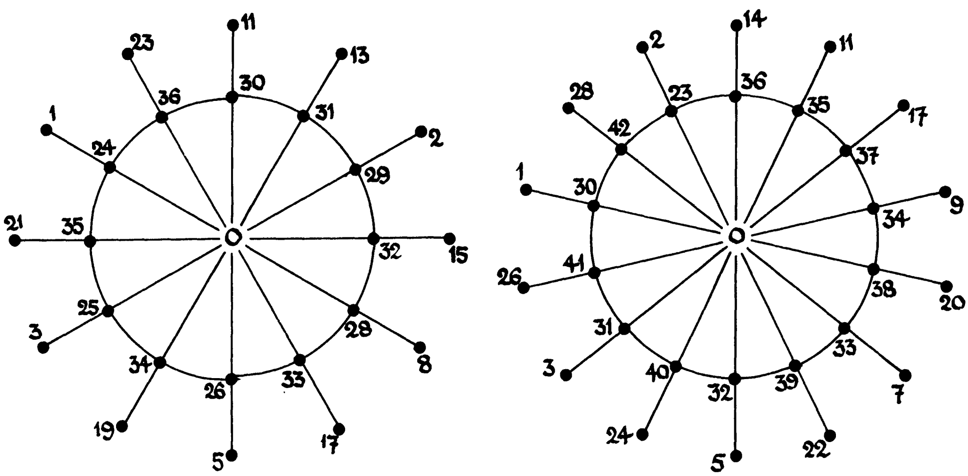


FIG. 1. Two graceful helms.

Peter Walker (1983) has the following result for Ogilvy's problem [1970, 388] on complex iterated radicals. The sequence $z_{n+1} = \sqrt{z_n + c}$ converges in all cases with $c < -1/4$, or with $\operatorname{Im} c > 0$, provided all square roots have positive imaginary part. For $\operatorname{Im} c < 0$ there are corresponding results when all square roots have negative imaginary part.

T. C. Brown [1971, 886] asked if there is a sequence on four symbols in which no two adjacent segments are permutations of one another, and discussed other aspects of Thue (Morse-Hedlund) sequences (Guy (1981), Problem E21). Some recent papers in this area are by Bean, Ehrenfeucht & McNulty (1979) and by Shelton (& Rao) (1981, 1982).

If S, C are noncoincident planar convex sets whose interiors intersect, define $\alpha(S, C)$ to be the number of connected components in the intersection of the boundaries of S and C, then Peterson [1972, 505] made the following three conjectures: (a) ¿ if $\alpha(S, S')$ is even or infinite for every S' congruent to S, then S has constant width? (b) ¿ if $\alpha(S, C)$ is even or infinite for every circular disc C of diameter w, then S has constant width w? (c) ¿ if C is a convex set of constant width w and $\alpha(S, C')$ is even or infinite for every C' congruent to C, then S has constant width w?

Goodey & Woodcock (1978) verified (b) and now Goodey (tbp) verifies all three conjectures by proving the following theorem: if S and C are planar convex sets such that $\alpha(S, C')$ is even or infinite for all translations C' of C, then S and C have equal width in every direction.

Detlef Seese (wrc) has shown that the crossing number and the rectilinear crossing number of a graph [1973, 52] are computable.

Earlier updating articles should have given more information on Lehmer's problem (Alter [1973, 192]): does $\phi(n)$, Euler's totient function, ever *properly* divide $n-1$? I quote from Guy (1981, Problem B37): Such an n must be a Carmichael number. Lehmer (1932) showed that n would have to be the product of at least seven distinct primes. Lieuwens (1970) proved the following theorems: if 3 divides n, then n is the product of more than 212 primes and $n > 5 \cdot 5 \times 10^{571}$; if the smallest prime factor of n is 5, n contains at least 11 prime factors; if the smallest prime factor of n is at least 7, then n is the product of at least 13 primes. This supersedes and corrects the work of Schuh (1944). Masao Kishore (1977) has shown that at least 13 primes are needed in any case and Cohen and Hagis (1980) have improved this to 14.

Klee [1977, 284] asked for a good algorithm to compute the d-measure of the union of n given ranges (hyperrectangles) in d-space. When $d \geqslant 3$, van Leeuwen & Wood (1981) improve Bentley's $O(n^{d-1}\ln n)$ algorithm (see Bentley & Wood (1980)) to one with $O(n^{d-1})$ steps.

For an excellent survey of the recent "dynamically increasing research activity in the theory of permanents," including the present status of Wang's conjectures [1978, 188; 1979, 119], see Minc (1983). Indeed, the present status is given for 20 conjectures and 10 problems, with a list of 11 additional conjectures and 3 further problems. There is a bibliography of 77 items published since Minc (1978) and 30 addenda to that earlier survey.

In answer to McCarty's question [1978, 578], Herzberg & Garner (1981) construct latin queen squares of any prime order, $p \geqslant 11$. For example, with (i, j)th entry $4i + 2j - 5 \pmod p$. John van Rees (1981) extends this to: latin queen squares of order n exist if gcd $(n, 210) = 1$, and (wrc) an analogous result for latin queen hypercubes in any number of dimensions. His two conjectures, (a) ¿ there are no toroidal latin queen squares of order n if gcd $(n, 210) > 1$? (b) ¿ there are no latin queen squares of order n if gcd $(n, 210) > 1$? are still open. He proves that toroidal latin queen squares of order n do not exist if gcd $(n, 6) > 1$, and mentions the problems: find a latin queen square of order 12, 14 or 15; find a toroidal latin queen square or order 25, 35 or 49.

In a remarkable paper, Tunnell (1983) has shown that the numbers less than a thousand and $\equiv 1, 2$ or $3 \pmod 8$ whose congruent or noncongruent status was unknown to Alter [1980, 43] and are missing from rows $a \equiv 1$, 2 or $3 \pmod 8$ in Table 1 [1981, 759] or Table 7 of Guy (1982, Problem D27), viz.

$a \equiv 1 \pmod 8$	569, 577, 593, 809, 857, 881, 897, 953
$a \equiv 2 \pmod 8$	282, 482, 706, 802, 898, 938
$a \equiv 3 \pmod 8$	627, 939

are all noncongruent. So only nine numbers less than a thousand have unknown status:

$a \equiv 5 \pmod 8$	573, 597, 677, 893, 917, 933, 965
$a \equiv 7 \pmod 8$	543, 623

Tunnell writes "It is a classical Diophantine problem to determine which integers are the area of some right triangle with rational sides. It is well known that D is the area of such a triangle if and only if the elliptic curve $y^2 = x^3 - D^2x$ has infinitely many rational points. Using results from the theory of modular forms of half-integral weight due to Shimura and Waldspurger, and the theorem of Coates-Wiles on elliptic curves with complex multiplication, we construct modular forms $\Sigma a(n)q^n$ of weight $3/2$ such that $a(n)$ nonzero implies that n is not the area of any rational

right triangle. The coefficients $a(n)$ are computed in terms of representations of n by a ternary quadratic form, and are conjecturally related to the order of the Tate-Shafarevitch group of the curve above."

For another paper on the postage stamp problem [1980, 206], see Mossige (1981).

In [1981, 760] I mentioned that a conjecture of Hamilton & Mullen [1980, 392] had been settled by a succession of correspondents. Jurai Bosák writes that the result was already in his book (1976): see also Laywine (1981).

Klee (1982) has settled the questions raised by Zowe [1980, 475]. In particular Zowe showed, for mathematical programming in infinite-dimensional vector spaces, that the Slater constraint qualification and a formally weaker one of Kurcyusz were equivalent in barrelled spaces, and asked if barrelledness is necessary. Klee has shown equivalence in a topological vector space E just if every barrel in E is a neighborhood of the origin. Thus, if E is locally convex, the two constraint qualifications are equivalent just if E is barrelled.

In his Zbl. review (444.05038) of Haggard's paper [1980, 654] A. T. White observes that Edmonds answers Question 1 (given a connected graph G and nonnegative integer n, is there a 2-cell imbedding of G whose geometric dual has n loops) in the affirmative when n is the number of edges in G, *not* for $n = 0$ as stated by Haggard.

Jacob Beard [1980, 744], and see Beard, Doyle & Mandelberg (1980), called a prime 3-**square-separable** if each pair of cubic residues which are not quadratic residues, is separated by a quadratic residue. For example, 43 is 3-s.-s. because (5 is a primitive root and) $5^3 \equiv 39$, $5^9 \equiv 22$, $5^{15} \equiv 8$, $5^{21} \equiv 42$, $5^{27} \equiv 27$, $5^{33} \equiv 2$, $5^{39} \equiv 32$ are separated by the quadratic residues (in parentheses) in

$$\ldots(1)2\ (4)\ 8\ (9)\ 22\ (25)\ 27\ (31)\ 32\ (36)\ 39\ (40)\ 42\ldots.$$

I asked if 7 is the only 3-s.-s. prime, so the answer is "no"! However, it seems likely that 7, 19, 37 and 43 are the only 3-s.-s. primes. Only a finite amount of calculation is needed to check this since the Lehmers (1962) with Mills (1963), Selfridge (1962) and Brillhart (1964) (and see Bierstedt & Mills (1963)) have results of the kind: "except for a finite number of primes, two consecutive cubic residues no larger than 77, 78 must occur." This needs extension to "cubic residues which are not quadratic residues," but presumably this is straightforward. See the Lehmers & Shanks (1970) and compare the argument: for a prime greater than 5, at least one of 2, 5 and 10 is a quadratic residue, so at least one of the pairs 1, 2 or 4, 5 or 9, 10 are quadratic residues. Note also the Lehmer-Mills-Selfridge result that, except for a finite number of primes, *three* consecutive cubic residues no larger than 23532, 23533, 23534 must occur, and see Sahib Singh (1977).

The corresponding situation for *quintic* residues seems to be that only 11, 31, 41, 61, 71 and 101 are 5-s.-s., but that needs more effort to check. The Lehmer-Mills bound for consecutive quintic residues is 7888, 7889. It seems reasonable to conjecture that there are only finitely many primes which are s-square-separable for each odd s. The Brillhart-Lehmer bound for consecutive septic (?) residues is 1649375, 1649376. The first few s-square-separable primes for $7 \leqslant s \leqslant 17$ are:

$s = 7$	29, 43, 113, 127	$s = 9$	19, 37, 73, 109, 181
$s = 11$	23, 67, 89, 199	$s = 13$	53, 79, 157
$s = 15$	31, 61, 151, 181, 211, 241, 271	$s = 17$	103, 137, 307

In his review (Zbl.463.10004), D. Wolke notes that some of the problems mentioned by Kenneth H. Rosen [1981, 276] were treated by E. Wirsing (1980). See also a further paper of Elliott (1976).

Suppose that $S = \{s_1, s_2, \ldots, s_n\}$ is a sequence of symbols, not necessarily distinct. A **rind** of S is a permutation $T = \{s_{\pi(1)}, s_{\pi(2)}, \ldots, s_{\pi(n)}\}$ of a special kind in which $\pi(i+1)$ is either the Least

or the gReatest of the set $\{1,2,\ldots,n\} \setminus \{\pi(1),\pi(2),\ldots,\pi(i)\}$, i.e., T is formed by peeling off the Leftmost or Rightmost symbol remaining in the original sequence S. Let $\rho(S)$ be the number of distinct rinds of S. It is easy to see that $\rho(S) \leqslant 2^{n-1}$ with equality just if the s_i are distinct. Göbel [1982, 113] conjectured that $\rho(S)$ is a maximum when the symbols in S are **grouped**, i.e., when like symbols occur contiguously. G. W. Peck (wrc) and P. J. Slater (wrc) notice that the conjecture is false; for example $\rho(AABBCCC) = 46$, whereas $\rho(ABBACCC) = 49$. In fact, if $n = 4k$ and each of the symbols $1,2,\ldots,2k$ occurs twice, Peck observes that $\{1,2,\ldots,k,1,2,\ldots,k,k+1,k+2,\ldots,2k,k+1,k+2,\ldots,2k\}$ has more rinds than the corresponding grouped sequence. However, this is still not best possible: $\rho(12123434) = 104$, whereas $\rho(12213443) = 110$. It seems surprisingly difficult to say anything that's both true and nontrivial.

The exhortation [1983, 35] not to try to solve certain problems may have been effective, though more earlier attempts have come to light. In connexion with Problem 1, on the distinctness of the Markov numbers, Schönheim sends a yellowing offprint of his 1956 paper.

I believe that it may be important to distinguish between Problems 2 and 3, i.e., between Problems E16 and E17 in Guy (1981). In the Collatz, '$3x + 1$' type of problem there is no (unique) inverse function, and the logical status may be different from the 'permutation sequences' which I attribute to Conway. I have given a brief partial history of the Collatz sequence [1983, 39] and Lagarias (tbp) is preparing the definitive article on the subject. Masai Yamada (1980) claimed to show that the sequence always collapsed to 1, but see Lagarias's review. Enrico Federighi sends the sequence $a_{n+1} = a_n/2$ (a_n even), $= a_n/3$ (a_n an odd multiple of 3), $= 7a_n + 1$ (otherwise). Does the 7 make it easier to prove anything? Are the cycles containing 1 and 19 the only ones?

As stated [1983, 40], Conway has shown that the premutation sequences belong to a class of problems which contains undecidable questions; but that is not to say you may not be able to decide a particular case. Such sequences appeared in problems of Klamkin (1963) and (a more tractable example; see Bergman (1964)) of Klamkin & Tritter (1963). Such problems were considered by John Conway, Alan Tritter and Mike Guy during Mike's undergraduate years (1960–63). See also comments by Shanks (1965) and Atkin (1966).

That the answer to Hartwig's question [1983, 120] about finding an invertible linear combination of matrices was "no for $n \geqslant 3$" was discovered in rapid succession by Stephen Boyack (83 : 02 : 10), Don Coppersmith, Victor Miller & Gadiel Seroussi (83 : 02 : 17 and 18), John H. Smith (83 : 02 : 22), Walter Sizer (83 : 02 : 23), Man-Duen Choi (83 : 03 : 09), Michael Atkinson (83 : 03 : 10), Jean-Marie Monier (83 : 03 : 21) and presumably many others. One of the simplest examples is

$$\begin{bmatrix} Q_1 & Q_3 \\ Q_2 & Q_4 \end{bmatrix} \text{ where } Q_1 = Q_4 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} Q_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} Q_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

Hartwig replied to Coppersmith, Miller & Seroussi:

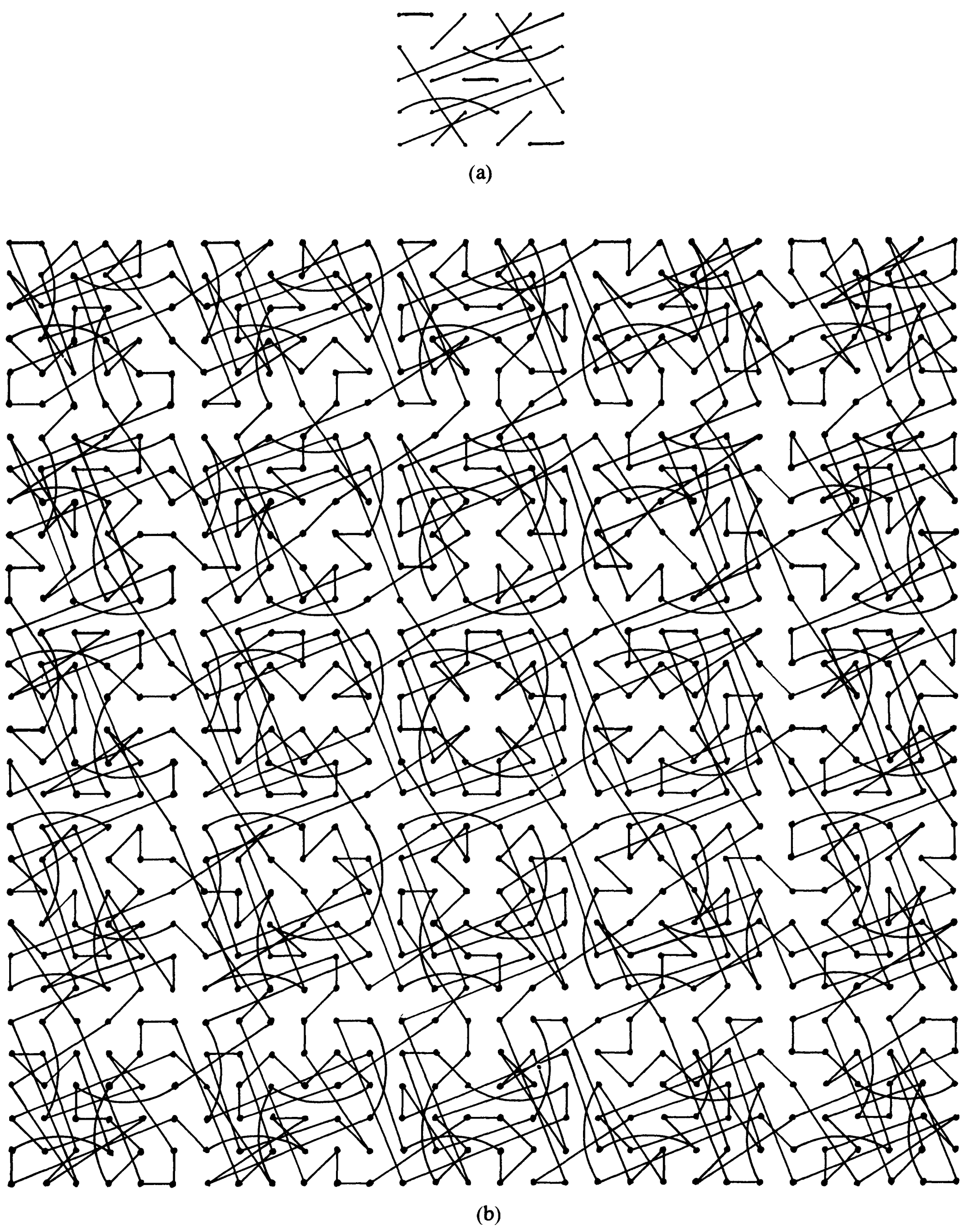
"The skew symmetric trick is very pretty and was found by several authors. It won't work for a division ring, however, as

$$U = \begin{bmatrix} 0 & \alpha & \beta \\ -\alpha & 0 & \gamma \\ -\beta & -\gamma & 0 \end{bmatrix}$$

can be invertible for $\beta\alpha^{-1}\gamma \neq \gamma\alpha^{-1}\beta$. Then $U\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. You are right, my question should have been: *when* do such scalars α_i exist for a field, a division ring, etc? The real problem is the splitting problem, which *does* hold for a field, but should be false for the integers."

Engel's "Celtic Art" problem [1983, 122] stimulated John P. Robertson to use a computer to find the 42 essentially different 3×4 matchings which lead to Hamilton cycles. There are two

different 2×3 matchings, each of which leads to a pattern with 4-fold rotational symmetry. These were depicted earlier [1983, 125]. Of the eight 2×5 matchings which generate Hamilton cycles, all but one have such symmetry. But none of the 42 (3×4)-generated patterns is symmetrical. Engel had earlier noted that this rotational symmetry only seems to occur if mn is singly even, not when mn is a multiple of 4. Is this a theorem, or haven't we searched far enough? [Stop press! Robertson writes that he has a proof.] Robertson has found symmetrical patterns generated by 3×10, 5×6 and 6×7 matchings, but none of the seven 4×5 matchings that he found leads to a symmetrical Hamilton circuit. Figure 2(a) is a 5×6 matching which gives the



(a)

(b)

FIG. 2. The matching (a) generates the Hamilton circuit (b).

symmetrical pattern shown in Figure 2(b). A reference omitted from the original article is Engel (1980).

Emeritus Professor Lawrence Ringenberg reminds us that snowplow problems [1983, 199] go back at least 32 years; see Klamkin (1951) and Ringenberg (1952).

My indebtedness to numerous correspondents is obvious. So too, I hope, is the usefulness of a clearing-house for information on problems which are not always quite as unsolved as we thought.

References

A. O. L. Atkin, Comment on Problem 63-13, SIAM Rev. 8 (1966) 234–236.

Jacqueline Ayel & Odile Favaron, The helms are graceful, #82T-05-202, Abstracts Amer. Math. Soc., 3(1982) 253.

Dwight R. Bean, A Ehrenfeucht & George F. McNulty, Avoidable patterns in strings of symbols. Pacific J. Math., 84 (1979) 261–269; MR 81i: 20075.

Jacob T. Beard, J. Kevin Doyle & Kenneth I. Mandelberg, Square-separable primes and unitary perfect polynomials, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) 68 (1980) 397–401.

J. L. Bentley & D. Wood, An optimal worst-case algorithm for reporting intersections of rectangles, IEEE Trans. Computers, C-29 (1980) 571–577; MR 81f: 68053.

George Bergman, Solution to Problem 5109, this MONTHLY, 71 (1964) 569–570.

R. G. Bierstedt & W. H. Mills, On the bound for a pair of consecutive quartic residues of a prime, Proc. Amer. Math. Soc., 14 (1963) 628–632; MR 27#4787.

Juraj Bosák, Latinské štvorce (latin squares), Mladá fronta, Prague, 1976.

J. Brillhart & D. H. & Emma Lehmer, Bounds for pairs of consecutive seventh and higher power residues, Math. Comput., 18 (1964) 397–407; MR 29#2214.

Graeme L. Cohen & Peter Hagis, On the number of prime factors of n if $\phi(n)|(n-1)$, Nieuw Arch. Wisk. (3) 28 (1980) 177–185; MR 81j: 10002

P. D. T. A. Elliott, On two conjectures of Kátai, Acta Arith., 30 (1976) 341–365; MR 54#5137.

Doug Engel, Wallpaper rings, J. Recreational Math., 13 (1980–81) 7–9.

K. J. Falconer, On the equireciprocal point problem, Geom. Dedicata, 14 (1983) 113–126.

Solomon Golomb, Normed division domains, this MONTHLY, 88 (1981) 680–686.

Paul Goodey, Intersections of planar convex curves (tbp).

Richard K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981.

Alfred P. Hallstrom, Equichordal and equireciprocal points, Bogazici Univ. J. Sci., 2 (1974) 83–88.

Agnes M. Herzberg & Cyril W. L. Garner, Latin queen squares, Utilitas Math., 20 (1981) 143–154.

Charlotte Huang, Anton Kotzig & Alexander Rosa, Further results on tree labellings, Utilitas Math., 21C (1982) 31–48.

Masao Kishore, On the number of distinct prime factors of n for which $\phi(n)|n-1$, Nieuw Arch. Wisk. (3) 25 (1977) 48–53; MR 56#2904.

M. S. Klamkin, Problem E963, this MONTHLY, 58 (1951) 260.

M. S. Klamkin, An infinite permutation, Problem 63-13, SIAM Rev., 5 (1963) 275–276.

M. S. Klamkin & A. L. Tritter, Problem 5109, this MONTHLY, 70 (1963) 572–573.

Victor Klee, A note on convex cones and constraint qualifications in infinite-dimensional vector spaces, J. Optimization Theory Appl., 37 (1982) 277–284.

Jeffery C. Lagarias, The "$3x + 1$" problem and its generalizations (tbp).

Charles Laywine, An expression for the number of equivalence classes of latin squares under row and column permutations, J. Combin. Theory Ser. A, 30 (1981) 317–320; MR 82j: 05029.

Jan van Leeuwen & Derick Wood, The measure problem for rectangular ranges in d-space, J. Algorithms, 2 (1981) 282–300; MR 82k: 68018.

D. H. Lehmer, On Euler's totient function, Bull. Amer. Math. Soc., 38 (1932) 745–751.

D. H. & Emma Lehmer, On runs of residues, Proc. Amer. Math. Soc., 13 (1962) 102–106; MR 25#2025.

D. H. & Emma Lehmer & W. H. Mills, Pairs of consecutive power residues, Canad. J. Math., 15 (1963) 172–177; MR 26#3660.

D. H. & Emma Lehmer, W. H. Mills & J. L. Selfridge, Machine proof of a theorem on cubic residues, Math. Comput., 16 (1962) 407–415; MR 28#5578.

D. H. & Emma Lehmer & Daniel Shanks, Integer sequences having prescribed quadratic character, Math. Comput., 24 (1970) 433–451; MR 42#5889.

E. Lieuwens, Do there exist composite numbers M for which $k\phi(M) = M - 1$ holds? Nieuw Arch. Wisk. (3) 18 (1970) 165–169; MR 42#1750.

P. Masai & A. Valette, A lower bound for a counterexample to Carmichael's conjecture, Boll. Unione Mat. Ital. (6) A1 (1982) 313–316.
Henryk Minc, Permanents, Encyclopedia of Mathematics and its Applications, vol. 6, Addison-Wesley, Reading, 1978; MR 80d : 15009.
Henryk Minc, Theory of permanents 1978–1981, Lin. Multilin. Alg., 12 (1983) 227–263.
Svein Mossige, Algorithms for computing the h-range of the postage stamp problem, Math. Comput., 36 (1981) 575–582; MR 82e : 10095.
L. A. Ringenberg, Solution to Problem E963, this MONTHLY, 59 (1952) 42.
Ioan Schönheim, Determinarea unei soluţii a ecuaţii $\Sigma_{j=1}^{N} x_j^2 = N\Pi_{j=1}^{N} x_j$ pentru orice întreg $N > 2$, x_j fiind numere întregi prime două câte două, Acad. R. P. Romîne Fil. Cluj. Stud. Cerc. Mat. Fiz., 7 (1956) no. 1-4, 59–63; MR 20#2302.
Detlef Seese, The crossing number of a graph is computable (wrc).
Daniel Shanks, Comment on Problem 63-13, SIAM Rev., 7 (1965) 285–286.
Robert O. Shelton, Aperiodic words on three symbols, J. Reine Angew. Math., 321 (1981) 195-209; II *ibid.*, 327 (1981) 1-11; III (with Raj P. Soni) *ibid.*, 330 (1982) 44–52; MR 82m : 05004.
Fred. Schuh, Can $n - 1$ be divisible by $\phi(n)$ when n is composite? Mathematica, Zutphen B., 12 (1944) 102–107; MR 7, 413f.
Sahib Singh, Bounds for the solutions of a Diophantine equation in prime Galois fields, Indian J. Pure Appl. Math., 8 (1977) 1428–1430; MR 80m : 10010.
Jerold B. Tunnell, A classical diophantine problem and modular forms of weight 3/2, J. Number Theory, 16 (1983).
Alain Valette, Fonction d'Euler et conjecture de Carmichael, Math. & Pédagog. no. 32 (1981) 13–18.
G. H. J. van Rees, On latin queen squares, Congr. Numer. 31, Proc. 10th Manitoba Conf. Numer. Math. Comput. 1980, vol. 2 (1981) 267–273; MR 83b : 05027.
Peter L. Walker, Complex iterated radicals, Math. Gaz., 67 (1983).
E. Wirsing, Additive functions with restricted growth on the numbers of the form $p + 1$, Acta Arith., 37 (1980) 345–357; Zbl. 445. 10042; MR 82b : 10056.
Masai Yamada, A convergence proof about an integral sequence, Fibonacci Quart., 18 (1980) 231–242; but see J. C. Lagarias, MR 82d : 10026.

NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

APPLICATIONS OF A RESULT ON SPHERICAL INTEGRATION TO THE THEORY OF CONVEX SETS

K. J. FALCONER
School of Mathematics, University of Bristol, University Walk, Bristol, England BS8 1TW

This note relates five results on the geometry of convex bodies by demonstrating that they are almost immediate consequences of a theorem on the integration of functions over Ω, the surface of the unit sphere in n-dimensional Euclidean space. This basic theorem is a corollary to the classical Funk-Hecke theorem concerning n-dimensional spherical harmonic functions. The article by Seeley [4] provides a readable account of spherical harmonics, and more detailed theory of such functions is given in Erdélyi [2]. Essentially the spherical harmonic functions $S_m(\theta)$ may be thought of as the higher dimensional analogues of cos mt and sin mt in Fourier series. If $f(\theta)$ is a reasonably behaved function on Ω, then we may write $f(\theta) = \Sigma_0^\infty c_m S_m(\theta)$ where $S_m(\theta)$ is some spherical harmonic of degree m.

If θ, ψ are two vectors in the $(n-1)$-sphere, let $\theta \cdot \psi$ denote their dot product. The Funk-Hecke theorem states: *If $g(x)$ is an L^1 function defined on the real interval $[-1, 1]$, then*

$$\int g(\theta \cdot \psi) S_m(\theta)\, d\theta = \lambda_m S_m(\psi) \tag{1}$$

for all $\psi \in \Omega$, *where*

$$\lambda_m = a_m \int_{-1}^{1} g(x) \frac{d^m}{dx^m} (1 - x^2)^{m+(1/2)n-3/2}\, dx \tag{2}$$

and a_m *is a nonzero number.*

Suppose $f(\theta)$ is a continuous function such that for all $\psi \in \Omega$, $\int g(\theta \cdot \psi) f(\theta)\, d\theta = 0$. We may obtain a form which is more useful for our purposes by multiplying equation (1) by $f(\psi)$ and integrating over Ω with respect to ψ. We get, after changing the order of integration,

$$\lambda_m \int f(\psi) S_m(\psi)\, d\psi = 0$$

for every mth degree spherical harmonic $S_m(\psi)$. Thus, provided that $\lambda_m \neq 0$,

$$\int f(\psi) S_m(\psi)\, d\psi = 0.$$

Then using the completeness of the spherical harmonics and the fact that mth degree spherical harmonics are odd or even according as to whether m is odd or even we obtain the following form:

Let $g(x)$ *be an* L^1 *function on* $[-1,1]$ *and let* λ_m *be as in* (2). *Suppose that* $f(\theta)$ *is a continuous function on* Ω *such that for all* $\psi \in \Omega$

$$\int g(\theta \cdot \psi) f(\theta)\, d\theta = 0.$$

Then

(A) *if* $\lambda_m \neq 0$ *for all* m, $f(\theta) \equiv 0$;
(B) *if* $\lambda_m \neq 0$ *for all odd* m, $f(\theta)$ *is an even function*;
(C) *if* $\lambda_m \neq 0$ *for all even* m, $f(\theta)$ *is an odd function.*

Finally for applications to convex sets we use the following special cases of the preceding form of the Funk-Hecke theorem.

(i) *Let* $f(\theta)$ *be a continuous function on* Ω *such that* $\int_{\theta\cdot\psi \geqslant 0} f(\theta)|\theta \cdot \psi|^k\, d\theta = 0$ *for all* $\psi \in \Omega$ *for some nonnegative integer* k. *Then if* k *is odd,* $f(\theta)$ *is an odd function, and if* k *is even,* $f(\theta)$ *is an even function.*

(ii) *Let* $f(\theta)$ *be a continuous function on* Ω *such that* $\int_{\theta\cdot\psi=0} f(\theta)\, d\theta = 0$ *for all* $\psi \in \Omega$ *(integration being with respect to the obvious invariant measure on the set* $\{\theta : \theta \in \Omega$ *and* $\theta \cdot \psi = 0\}$*). Then* $f(\theta)$ *is an odd function.*

(i) may be obtained by putting $g(x) = x^k$ for $0 \leqslant x \leqslant 1$ and $g(x) = 0$ otherwise. We may use repeated integration by parts (and Leibniz's theorem) to see that we have the situation of (B) above if k is even and of (C) above if k is odd. For case (ii) we take $g(x)$ to be a "delta" function (using the usual method of approximate identities) giving

$$\lambda_m = a_m \left[\frac{d^m}{dx^m} (1 - x^2)^{m+(1/2)n-3/2} \right]_{x=0}$$

and as this is nonzero for even m, (C) above yields the result.

This concludes our preparatory work with the Funk-Hecke theorem.

We now deduce various results on convex bodies; all convex sets to be considered are assumed closed and bounded and not contained in any hyperplane. The first result gives a sufficient condition for a convex body to be centrally symmetric; such results are important when defining measures of symmetry of convex bodies. (A measure of symmetry is a function that takes value 1 on centrally symmetric sets and values strictly less than 1 on other convex sets, see Grünbaum [3].)

THEOREM 1. *Let X be a convex body in R^n, and let P be a point such that every hyperplane through P divides X into two parts of equal n-dimensional volume. Then X is centrally symmetric with P as center.*

Proof. Certainly P is an interior point of X. Let $f(\theta)$ be the defining function of X with respect to P, so that $(\theta, f(\theta))$ gives the boundary of X in polar coordinates with P as origin. Considering the portion of X cut off by the hyperplane through P and perpendicular to the unit vector ψ, we must have that

$$\frac{1}{n}\int_{\theta\cdot\psi\geqslant 0} f(\theta)^n \, d\theta = \frac{1}{2}V$$

for all ψ, where V is the volume of X. Thus

$$\int_{\theta\cdot\psi\geqslant 0}\left[f(\theta)^n - f(-\theta)^n\right] d\theta = 0$$

for all ψ. By (i) above $f(\theta)^n - f(-\theta)^n$ is an even function, so as it is certainly an odd function as well, $f(\theta) = f(-\theta)$ for all θ.

Next we show that a centrally symmetric convex set is determined by the areas of its sections through the center of symmetry.

THEOREM 2. *Let X and Y be centrally symmetric convex bodies in R^n such that the $(n-1)$-dimensional areas of parallel sections through the centers of the bodies are equal. Then X and Y are translates of each other.*

Proof. Let $f(\theta)$ and $g(\theta)$ be the defining functions of the bodies with respect to their centers. Thus

$$\frac{1}{n-1}\int_{\theta\cdot\psi=0} f(\theta)^{n-1} \, d\theta = \frac{1}{n-1}\int_{\theta\cdot\psi=0} g(\theta)^{n-1} \, d\theta$$

for all ψ. (ii) implies that $f(\theta)^{n-1} - g(\theta)^{n-1}$ is an odd function, so that as it is also an even function (from the central symmetry of X and Y), $f(\theta) = g(\theta)$ for all θ.

Ulam [5] p. 38 mentions the following problem, as yet unsolved: Let X be a convex body in R^3 made of a material of uniform density ρ with $0 < \rho < 1$. Suppose that X will float in equilibrium in any orientation in a liquid of density 1, must X be spherical? We show that this is true in a special case:

THEOREM 3. *Let X be a centrally symmetric convex body in R^3 of uniform density $\frac{1}{2}$ that will float in equilibrium in any orientation in a liquid of density* 1. *Then X is spherical.*

Proof. Let X be defined by the function $f(\theta)$, taking the center of symmetry as origin. When X floats with the vector ψ pointing downwards, the parts of the surface of X with $\theta \cdot \psi > 0$ are submerged. By the energy principle, if equilibrium is to be possible in all orientations, the center of gravity of the part of X below the surface level of the liquid must be at constant depth. That is

$$\int_{\theta\cdot\psi\geqslant 0} f(\theta)^4 |\theta\cdot\psi| \, d\theta = c$$

for all ψ, where c is constant. Hence for some number k

$$\int_{\theta\cdot\psi\geqslant 0}\left[f(\theta)^4 - k\right]|\theta\cdot\psi| \, d\theta = 0$$

for all ψ. By (i) this implies $f(\theta)^4 - k = 0$ for all θ, making X spherical.

Next we show that a symmetric convex body is determined by the perimeter of its shadow in every direction.

THEOREM 4. *Let X and Y be centrally symmetric convex bodies in R^n ($n \geqslant 3$) such that for every hyperplane H the $(n-2)$-dimensional perimeter areas of the projections of X and Y onto H are equal. Then X and Y are translates of each other.*

Proof. Let X and Y have support functions $h(\theta)$ and $k(\theta)$ with respect to their centers. Using Cauchy's surface area formula (see for example Eggleston [1] p. 89) we get, considering projections onto the hyperplane perpendicular to ψ,

$$\int_{\theta\cdot\psi=0} h(\theta)\, d\theta = \int_{\theta\cdot\psi=0} k(\theta)\, d\theta$$

for all ψ. We deduce from (ii) that $h(\theta) = k(\theta)$ for all θ so that X and Y are translates.

Finally, if X is a body of constant width in R^3, it follows from the plane version of Cauchy's surface area formula that all plane projections of X (which must also be sets of constant width) have equal perimeters. Our last result shows that the converse of this is also true.

THEOREM 5. *Let X be a convex body in R^3 such that every plane projection of X has perimeter of length k. Then X is of constant width k/π.*

Proof. Let X have width $w(\theta)$ in direction θ. From Cauchy's surface area formula

$$\int_{\theta\cdot\psi=0} w(\theta)\, d\theta = k$$

for all ψ, giving that

$$\int_{\theta\cdot\psi=0} [w(\theta) - k/\pi]\, d\theta = 0$$

for all ψ. By (ii) we have $w(\theta) = k/\pi$ for all θ, as required.

References

1. H. G. Eggleston, Convexity, Cambridge University Press, 1969.
2. A. Erdélyi, et al., Higher transcendental functions, vol. 2, Bateman manuscript project, New York, 1953.
3. B. Grünbaum, Measures of symmetry for convex sets, Convexity, Proceedings of symposia in Pure Mathematics, vol. 7, American Mathematical Society, Providence, 1963.
4. R. T. Seeley, Spherical harmonics, this MONTHLY, 73 (1966), Slaught memorial supplement, 115–121.
5. S. M. Ulam, A collection of mathematical problems, Interscience, New York, 1960.

ITERATING THE DERIVED SET FUNCTION

DENIS HIGGS
Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Cantor, in investigating the convergence of Fourier series, was led around 1870 to define the derived set (the set of all limit points) $d(A)$ of a subset A of the real line $\mathbb{R}$. He was also led to consider the iteration of the derived set function d, thereby coming to define ordinal numbers α in order to define the αth iterate d^α of d (see [1] and [2, III]). An interesting fact, which does not seem to have been noticed hitherto, is that d^α is itself the derived set function for some topology on $\mathbb{R}$. Indeed, for very many topological spaces S, it turns out that d^α is the derived set function for a new topology on S.

This note first characterizes the topological spaces for which all the d^α functions are derived set functions. Then the unfortunate extent to which iteration fails to relate well to various standard topological notions is noted, and the "ultimate iterate" of d (already considered by Cantor, for the case $S = \mathbb{R}$) is defined, along with a related concept.

The derived set $d(A)$ of a subset A of a topological space S is the set of all limit points in S of A, so that a point x is in $d(A)$ if and only if every open set containing x meets A in a point $y \neq x$. Let S be any set and let $\mathcal{P}(S)$ denote the power set of S. A function $d: \mathcal{P}(S) \to \mathcal{P}(S)$ which is the derived set function for some topology on S will be called a *derived set function* on S. As is well known, a topology on S can always be recovered from the associated derived set function d since the closure of a subset A of S is just $A \cup d(A)$. For this reason we sometimes refer to a topological space as a pair (S, d) where d is a derived set function on S. Derived set functions may be characterized as follows (Dugundji [3, p. 73]; similar characterizations are given by Harvey [4], Thron [7, p. 53], and others):

PROPOSITION 1. *A function $d: \mathcal{P}(S) \to \mathcal{P}(S)$ is a derived set function on S if and only if*

(i) $d(\varnothing) = \varnothing$ *and* $d(A \cup B) = d(A) \cup d(B)$ *for all* $A, B \in \mathcal{P}(S)$,
(ii) $x \notin d(x)$ *for all* $x \in S$,
(iii) $d^2(A) \subseteq A \cup d(A)$ *for all* $A \in \mathcal{P}(S)$.

(Here $d(x)$ should really be $d(\{x\})$, and $d^2(A)$ means $d(d(A))$.)

The αth iterate $d^\alpha: \mathcal{P}(S) \to \mathcal{P}(S)$ of a derived set function d on S is defined by transfinite recursion on α as follows:

$$d^1(A) = d(A), \qquad d^{\alpha+1}(A) = d(d^\alpha(A)), \qquad d^\lambda(A) = \bigcap_{\alpha<\lambda} d^\alpha(A),$$

where A is any member of $\mathcal{P}(S)$, α denotes a nonzero ordinal, and λ denotes a limit ordinal.

In order to determine for which derived set functions d every d^α is also a derived set function, consider first the case $\alpha = 2$. In view of condition (ii) above, a necessary condition for d^2 to be a derived set function is that $x \notin d^2(x)$ for all $x \in S$. Now $x \notin d^2(x)$ if and only if $d^2(x) \subseteq d(x)$, that is, if and only if $d(x)$ is closed. So our necessary condition is that $d(x)$ is closed for all $x \in S$ and this is just the T_D separation axiom of Aull and Thron (see [7, p. 92]; T_D is not strong, being intermediate to T_0 and T_1). From [7, p. 93] we have:

PROPOSITION 2. *A function $d: \mathcal{P}(S) \to \mathcal{P}(S)$ is a T_D derived set function (that is, the derived set function for a T_D topology) if and only if* (i) *and* (ii) *above hold and also*

(iii)$_D$ $d^2(A) \subseteq d(A)$ *for all* $A \in \mathcal{P}(S)$.

We now use this result to show that if d is a T_D derived set function, then so is every d^α. For functions $f: \mathcal{P}(S) \to \mathcal{P}(S)$ the notation $f_1 \leqslant f_2$ will be used to mean that $f_1(A) \subseteq f_2(A)$ for all $A \in \mathcal{P}(S)$.

THEOREM. *Let d be a derived set function on a set S. Then d^α is a T_D derived set function on S for all nonzero ordinals α. Moreover, $d^\beta \leqslant d^\alpha$ for $\alpha \leqslant \beta$, and $d^{\alpha+\beta} = d^\beta d^\alpha$, $d^{\alpha\beta} = (d^\alpha)^\beta$ for all α and β.*

Proof. Let $P(\alpha)$ be the assertion that $d^{\alpha+1} \leqslant d^\alpha$. Then $P(1)$ holds since d is T_D, $P(\alpha)$ clearly implies $P(\alpha+1)$, and if $P(\alpha)$ holds for all $\alpha < \lambda$, then for any $A \in \mathcal{P}(S)$ we have

$$d^{\lambda+1}(A) = d(d^\lambda(A)) = d\Big(\bigcap_{\alpha<\lambda} d^\alpha(A)\Big) \subseteq d(d^\alpha(A)) = d^{\alpha+1}(A) \subseteq d^\alpha(A)$$

for all $\alpha < \lambda$, from which $P(\lambda)$ follows. Therefore $P(\alpha)$ holds for all α by transfinite induction on α. An even simpler induction on β now shows that $d^\beta \leqslant d^\alpha$ for $\alpha \leqslant \beta$.

Let $Q(\alpha)$ state that d^α satisfies (i) above. $Q(1)$ is certainly true and $Q(\alpha)$ clearly implies $Q(\alpha+1)$. Suppose $Q(\alpha)$ holds for all $\alpha < \lambda$. Then $Q(\lambda)$ states that

$$\bigcap_{\alpha<\lambda} d^\alpha(A) \cap d^\alpha(B) = \bigcap_{\alpha<\lambda} d^\alpha(A) \cap \bigcap_{\alpha<\lambda} d^\alpha(B)$$

and it is clear that the left-hand side is contained in the right. To show that the other containment

also holds, let x be in the right-hand side, so that $x \in d^{\alpha}(A) \cap d^{\beta}(B)$ for some $\alpha, \beta < \lambda$ where $\alpha \leqslant \beta$, say. Then $x \in d^{\alpha}(A) \cap d^{\alpha}(B)$ and hence x is in the left-hand side. So $Q(\lambda)$ holds, and we therefore have $Q(\alpha)$ for all α, that is, d^{α} satisfies condition (i) for all α.

Since $d^{\alpha}(x) \subseteq d(x)$, d^{α} satisfies (ii) for all α.

Let $R(\alpha)$ state that d^{α} satisfies $(\text{iii})_D$. Then $R(1)$ holds. If $R(\alpha)$ holds, then for any $A \in \mathcal{P}(S)$ we have

$$(d^{\alpha+1})^2(A) = d\big(d^{\alpha}\big(d^{\alpha+1}(A)\big)\big) \subseteq d\big(d^{\alpha}\big(d^{\alpha}(A)\big)\big) \subseteq d\big(d^{\alpha}(A)\big) = d^{\alpha+1}(A)$$

so that $R(\alpha+1)$ holds (the first inclusion here uses $P(\alpha)$ and the fact that $A \subseteq B$ implies $d^{\alpha}(A) \subseteq d^{\alpha}(B)$ which we have on account of $Q(\alpha)$). If $R(\alpha)$ holds for all $\alpha < \lambda$ then

$$(d^{\lambda})^2(A) = \bigcap_{\alpha<\lambda} d^{\alpha}\big(d^{\lambda}(A)\big) \subseteq \bigcap_{\alpha<\lambda} d^{\alpha}\big(d^{\alpha}(A)\big) \subseteq \bigcap_{\alpha<\lambda} d^{\alpha}(A) = d^{\lambda}(A)$$

for all $A \in \mathcal{P}(S)$ so that $R(\lambda)$ holds. Hence d^{α} satisfies condition $(\text{iii})_D$ for all α.

It only remains to show the two identities involving ordinal addition and multiplication and we omit the routine arguments.

An iterate d^{α} with $\alpha > 2$ of a derived set function d can itself be a derived set function without d being T_D. Indeed, the two-point indiscrete space provides a simple instance. It can be shown that the following three conditions are equivalent for any derived set function d: d^3 is a derived set function; $x \notin d^3(x)$ for all x; and $d^3(A) \subseteq d(A)$ for all A. I do not know if this generalizes.

Now we turn to the bad behaviour of iteration. To observe this it is already sufficient in most cases to look at d^2 where d is the usual derived set function on $\mathbb{R}$. First note that for $x \in \mathbb{R}$ and $A \in \mathcal{P}(\mathbb{R})$ we have $x \in d^2(A)$ if and only if $A \cap \{y: 1/(m+1) < |y-x| \leqslant 1/m\}$ is infinite for infinitely many $m \in \mathbb{N} = \{1,2,3,\dots\}$. It follows that a basis of d^2-neighbourhoods of x is obtained by taking the sets of the form $[x-(1/m), x+(1/m)] \setminus \{x_n : n \in \mathbb{N}\}$ where $m \in \mathbb{N}$ and x_n, $n \in \mathbb{N}$, is a sequence of points such that $x_n \neq x$ for each n and $\lim_{n\to\infty} x_n = x$. What is relevant here is that $\mathbb{R}$ is a locally compact metric space—thus a similar description of d^2 and d^2-neighbourhoods, and many of the remarks here, are valid for any such space.

How do various topological properties of $(\mathbb{R}, d)$ fare in passing to $(\mathbb{R}, d^2)$? It is easy to see from the above description of d^2-neighbourhoods that $(\mathbb{R}, d^2)$ is neither regular, nor locally compact, nor locally connected. $(\mathbb{R}, d^{\alpha})$ is Hausdorff for all α, this simply being a consequence of the fact that Hausdorffness is expansion-invariant—that is, is preserved by enlarging the topology (since $d^{\alpha} \leqslant d$, the d^{α}-topology is an expansion of the d-topology). Since nonconnectedness is expansion-invariant, it is more surprising that $(\mathbb{R}, d^{\alpha})$ is connected for all α, a fact which can perhaps best be seen to hold by showing, with the aid of the Cantor-Bendixson Theorem (see [**2**]), that $(\mathbb{R}, d^{\rho(d)})$ is connected, where $d^{\rho(d)}$ is the ultimate iterate of d defined below. However, this fact is somewhat special to $(\mathbb{R}, d)$ and in general connectedness is lost in passing from (S, d) to (S, d^2).

Iteration does not restrict nicely to subspaces. If T is a subset of a space (S, d), let d_T denote the derived set function on T with the subspace topology. Then for most T_D spaces, it is not the case that $(d_T)^{\alpha} = (d^{\alpha})_T$ for all T and α, a counterexample being given by taking the usual d on $\mathbb{R}$ with

$$T = \left\{ \frac{1}{m} + \frac{1}{n} : m, n \in \mathbb{N} \right\} \cup \{0\};$$

then $0 \in (d^2)_T(T)$ but $0 \notin (d_T)^2(T)$. It is easily verified that we always have $(d_T)^{\alpha} \leqslant (d^{\alpha})_T$, with equality holding if T is either open or closed. The T_D spaces in which $(d_T)^{\alpha} = (d^{\alpha})_T$ for all T and α turn out to coincide with those for which d is "hereditarily idempotent" (see the remarks on the ultimate iterate below; such spaces are discussed in [**6**]).

Iteration does not behave well on products. Let $d_1 \times d_2$ denote the derived set function on a product space $(S_1, d_1) \times (S_2, d_2)$. Then in general $(d_1 \times d_2)^{\alpha} \neq d_1^{\alpha} \times d_2^{\alpha}$: if we take the usual d

on $\mathbb{R}$ and let $A = \{(1/m, 1/n) \colon m, n \in \mathbb{N}\}$, then $(d \times d)^2(A) = \{(0,0)\}$ but $(d^2 \times d^2)(A) = \varnothing$.

Iteration does not behave well with respect to continuous functions. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function whose graph is shown in Fig. 1. Then f is continuous with respect to the usual d but not with respect to d^2, as can be seen from the fact that if

$$A = \bigcup_{n \in \mathbb{N}} \left[\frac{1}{2n}, \frac{1}{2n-1}\right],$$

then $0 \in d^2(A)$ but $f(0) \notin f(A) \cup d^2(f(A))$. Something can be salvaged here nevertheless: say that a function $f\colon S_1 \to S_2$ is *strictly* continuous with respect to derived set functions d_1, d_2 on S_1, S_2 respectively if $x \in d_1(A)$ implies $f(x) \in d_2(f(A))$ for all $x \in S_1$ and $A \in \mathcal{P}(S_1)$. It is easy to see that if f is strictly continuous with respect to d_1, d_2 and d_1, d_2 are T_D, then f is strictly continuous with respect to d_1^α, d_2^α for all α.

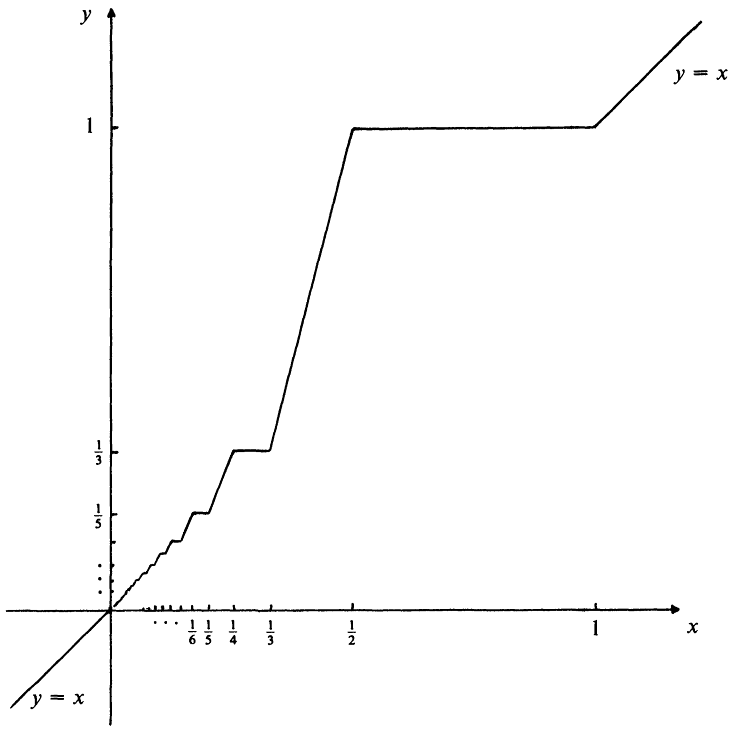


FIG. 1

Although $(\mathbb{R}, +)$ is a topological group for the (usual) d-topology, this is no longer true for the d^2-topology:

$$U = [-1, 1] \setminus \left\{\frac{1}{n} \colon n \in \mathbb{N}\right\}$$

is a d^2-neighbourhood of 0 but there is no d^2-neighbourhood V of 0 such that $V + V \subseteq U$, this being because $V + V \subseteq U$ implies that $V \times V$ is disjoint from each of the diagonal lines $x + y = 1/n$, an impossibility.

We close by mentioning the ultimate iterate $d^{\rho(d)}$ of a T_D derived set function d on a set S. Let $|\mathcal{P}(S)^{\mathcal{P}(S)}| = \kappa$—then since d^α decreases with increasing α and has at most κ distinct values, there must be some ordinal ρ no larger in cardinality than κ such that $d^\alpha = d^\rho$ for all $\alpha \geqslant \rho$. Let us call the smallest such ρ the *rank* $\rho(d)$ of d. The fact that every nonzero ordinal occurs as the rank of

some d is a consequence of the following result, due essentially to Hausdorff [**5**, p. 280].

PROPOSITION 3. *If an ordinal $\sigma > 1$ is given the order topology, then the rank of the associated d is the smallest ordinal ρ such that $\omega^{\rho} \geqslant \sigma$.*

In order to describe $d^{\rho(d)}$ without reference to iteration (and also to describe a related concept), we need the following definitions. Let (S, d) be any topological space. Then d is *idempotent* if $d^2 = d$ and *hereditarily* idempotent if $(d_T)^2 = d_T$ for all $T \subseteq S$. For $A \in \mathcal{P}(S)$, let $c(A)$ denote the closure of A and let $p(A)$ denote the largest dense-in-itself subset of A (a subset B of S is dense-in-itself if $d(B) \supseteq B$). For (S, d) a T_D space, $d^{\rho(d)}$ is clearly the largest idempotent derived set function $\leqslant d$. Also, as was essentially known to Cantor and as is in any case easily verified, we have $d^{\rho(\alpha)} = pc$ (where pc means "c first, then p").

In a certain sense, $d^{\rho(d)}$ is not quite ultimate: although idempotent, it is not hereditarily idempotent in general (for instance, it is not in the case of the usual d on $\mathbb{R}$). All the same, there does exist a largest hereditarily idempotent derived set function $\leqslant d$, namely cp. In fact, cp is a hereditarily idempotent derived set function on any T_0 space S. One way to see this is as follows. First note that the scattered subsets of S form an ideal $\mathbb{S}$ ($A \subseteq S$ is scattered if and only if it contains no nonempty dense-in-itself set). Hence $\mathbb{S}$ may be minimally adjoined to the stock of closed subsets of S by means of the "localization" construction described by Vaidyanathaswamy [**8**, p. 171 et seq.]. Now, as Vaidyanathaswamy proves, the derived set function of the augmented topology is just cp [**8**, p. 183]). Moreover, since $\mathbb{S}$ is an adherence-ideal in the terminology of [**8**] (see pp. 177 and 183), it follows from results of [**6**] that cp is hereditarily idempotent and is furthermore the largest hereditarily idempotent derived set function $\leqslant d$.

References

1. G. Cantor, Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen, Math. Ann., 5 (1872) 123–132.

2. ______, Über unendliche, lineare Punktmannigfaltigkeiten I, Math. Ann., 15 (1879) 1–7; II, ibid., 17 (1880) 355–358; III, ibid., 20 (1882) 113–121; IV, ibid., 21 (1883) 51–58; V, ibid., 21 (1883) 545–591; VI, ibid., 23 (1884) 453–488.

3. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.

4. F. R. Harvey, The derived set operator, this MONTHLY, 70 (1963) 1085–1086.

5. F. Hausdorff, Grundzuge der Mengenlehre, reprint, Chelsea, New York, 1949.

6. D. Higgs, Topological spaces which are also matroidal, preprint, 1981.

7. W. J. Thron, Topological Structures, Holt, Rinehart and Winston, New York, 1966.

8. R. Vaidyanathaswamy, Set Topology, 2nd ed., Chelsea, New York, 1960.

CROSS PRODUCTS OF VECTORS IN HIGHER DIMENSIONAL EUCLIDEAN SPACES

W. S. MASSEY

Department of Mathematics, Yale University, New Haven, CT 06520

Several times I have had students in undergraduate mathematics courses ask me the following question: Can one define a cross product of vectors in Euclidean n-space for $n > 3$ so that it will have properties similar to the usual cross product of vectors in 3-space? Of course the answer to this question will probably depend on which properties of the usual cross product one requires to hold in n-space; it is conceivable that there will be many different answers, depending on which properties are required to hold.

Fortunately the situation is not quite as chaotic as the foregoing sentences might suggest. If one requires only three basic properties of the cross product, properties which are explained in practically all undergraduate textbooks that discuss vector analysis, it turns out that a cross product of vectors exist only in 3-dimensional and 7-dimensional Euclidean space. To the best of the author's knowledge, the only textbook which contains a discussion of this fact is Hilton [**8**].

The purpose of this note is to give an explanation of this result which will be accessible to the average reader of this MONTHLY. We will actually give two theorems on this subject: the first uses purely algebraic techniques, and is based on a famous theorem proved by A. Hurwitz in 1898. The second theorem gives a stronger result; it depends on a deep theorem proved by J. F. Adams in 1958. At the end of the paper we discuss some other results in this area.

Our notation is standard: R^n denotes the real vector space consisting of n-tuples of real numbers,

$$x \cdot y = \sum x_i y_i$$

is the dot product of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, and

$$|x| = (x \cdot x)^{1/2}$$

denotes the norm or length of the vector x.

THEOREM I. *Assume $n \geqslant 3$ and a cross product is defined which assigns to any two vectors $v, w \in R^n$ a vector $v \times w \in R^n$ such that the following three properties hold*:

(a) *$v \times w$ is a bilinear function of v and w.*
(b) *The vector $v \times w$ is perpendicular to both v and w, i.e., $(v \times w) \cdot v = (v \times w) \cdot w = 0$.*
(c) *$|v \times w|^2 = |v|^2|w|^2 - (v \cdot w)^2$.*

Then $n = 3$ or 7.

REMARK. Note that condition (c) is the usual condition that the length of $v \times w$ shall be equal to the area of the parallelogram spanned by v and w.

Proof. The proof consists in showing that a cross product defined on R^n and having the three properties listed above implies the existence of a bilinear multiplication on R^{n+1} which has very special properties. We will consider R^{n+1} as an orthogonal direct sum:

$$R^{n+1} = R^1 \oplus R^n.$$

Thus an element of R^{n+1} consists of an ordered pair (a, v), where a is a real number and $v \in R^n$. The required product is defined by the following formula:

$$(a, v)(b, w) = (ab - v \cdot w, aw + bv + v \times w). \tag{1}$$

This multiplication is obviously bilinear, and $(1, 0)$ is a 2-sided unit. An easy computation using properties (b) and (c) shows that the norm of the product of two elements of R^{n+1} is given by the following formula:

$$|(a, v)(b, w)|^2 = |(a, v)|^2 |(b, w)|^2 \tag{2}$$

Now this is exactly the situation considered by A. Hurwitz [**4**] in 1898. Hurwitz proved that if we have a bilinear multiplication with a unit defined on R^q such that the norm of the product of two vectors is the product of the norms (condition (2) above), then q must be 1, 2, 4, or 8, and the multiplication is isomorphic to that of the real numbers, the complex numbers, the quaternions, or the octonions of Cayley and Graves. For a lucid exposition of a modern version of this theorem of Hurwitz, see Jacobson, [**5**, pp. 417–427].

Note that the uniquess assertion of Hurwitz's theorem shows that conditions (a), (b), and (c) of Theorem I characterize the cross products on R^3 and R^7 uniquely up to isomorphism.

The interested reader is referred to pp. 408–409 of a paper by E. Calabi [**2**] for a list of additional properties of the cross product in R^7 and an actual multiplication table for this cross product in terms of an orthonormal basis of R^7.

REMARK. The multiplication given by formula (1) has evidently been known for a long time. In 1942 B. Eckmann referred to it as "einer bekannten, elementaren Konstruktion" (see [**3**, p. 338]).

In our next theorem we show that we can significantly weaken conditions (a) and (c) of Theorem I without altering the conclusion.

THEOREM II. *Assume $n \geqslant 3$ and that a cross product is defined which assigns to any two vectors $v, w \in R^n$ a vector $v \times w$ such that the following three properties hold*:

(a) *$v \times w$ is a continuous function of the ordered pair (v, w).*
(b) *The vector $v \times w$ is perpendicular to both v and w, i.e., $(v \times w) \cdot v = (v \times w) \cdot w = 0$.*
(c) *If v and w are linearly independent, then $v \times w \neq 0$.*

Then $n = 3$ or 7.

Proof. For any vectors $v, w \in R^n$, let

$$A(v, w) = \left[|v|^2|w|^2 - (v \cdot w)^2\right]^{1/2}.$$

Then $A(v, w)$ is equal to the area of the parallelogram spanned by v and w; it is obviously a continuous function of the ordered pair (v, w). Using this area function, we define a function

$$f\colon R^n \times R^n \to R^n$$

by the formula

$$f(v, w) = \begin{cases} \dfrac{A(v, w)}{|v \times w|}(v \times w) & \text{if } v \times w \neq 0, \\ 0 & \text{if } v \times w = 0. \end{cases}$$

We assert that the function f thus defined is continuous. To prove this, it suffices to prove that if (v_k, w_k) is any infinite sequence of pairs of vectors such that

$$\lim_{k \to \infty} (v_k, w_k) = (v_0, w_0),$$

then

$$\lim_{k \to \infty} f(v_k, w_k) = f(v_0, w_0).$$

There are various cases to consider, depending on the two cases in the definition of $f(v, w)$, but the details are completely elementary.

Note that the function f thus defined satisfies the following two conditions:

$$f(v, w) \cdot v = f(v, w) \cdot w = 0, \tag{3}$$

$$|f(v, w)|^2 = |A(v, w)|^2 = |v|^2|w|^2 - (v \cdot w)^2. \tag{4}$$

Exactly as before, we may consider R^{n+1} as the direct sum $R^1 \oplus R^n$, and define a function

$$\mu\colon R^{n+1} \times R^{n+1} \to R^{n+1}$$

by the formula

$$\mu[(a, v), (b, w)] = (ab - v \cdot w, aw + bv + f(v, w)) \tag{5}$$

(compare with (1)). Then μ is obviously continuous and $(1, 0)$ is a 2-sided unit in the sense that

$$\mu[(1,0), (a, v)] = \mu[(a, v), (1,0)] = (a, v).$$

Finally, we have the analog of formula (2):

$$|\mu(x, y)|^2 = |x|^2|y|^2 \tag{6}$$

for any $x, y \in R^{n+1}$. Now let S^n denote the unit n-dimensional sphere:

$$S^n = \{x \in R^{n+1} \mid |x| = 1\}.$$

In view of formula (6), we see that if x and y belong to S^n, then $\mu(x, y) \in S^n$ also. Thus μ defines a continuous multiplication with a 2-sided unit on the n-sphere S^n.

This raises the following question: For what values of n does the n-sphere admit a continuous multiplication with a 2-sided unit? This was a famous problem for many years in algebraic

topology. It was finally resolved by Frank Adams in 1958 (see [1]). The answer is that such a continuous multiplication exists on S^n only in the cases $n = 1$, 3, and 7. Examples of such multiplications arise from the multiplication of complex numbers, quaternions, and the Cayley–Graves octonions respectively, restricted to the unit sphere.

Thus we see that by referring to this theorem of Adams we can complete the proof of Theorem II.

We will conclude this note by considering other possible ways to generalize the definition of the cross product to higher dimensional Euclidean spaces. Let v and w be vectors in R^3; recall the formula for the components of $v \times w$ in terms of the components of v and w. According to this formula, if

$$v = (v_1, v_2, v_3)$$
$$w = (w_1, w_2, w_3)$$

then the kth component of $v \times w$ is the determinant of the 2×2 submatrix of the matrix

$$\begin{bmatrix} v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

obtained by striking out the kth column. To be correct, we must multiply this determinant by $(-1)^k$.

Analogously, in R^n we can define a cross product $v_1 \times v_2 \times \cdots \times v_{n-1}$ of any ordered $(n-1)$ tuple of vectors by a similar process. Form a matrix whose successive rows are the vectors $v_1, v_2, \ldots, v_{n-1}$. The kth component of $v_1 \times v_2 \times \cdots \times v_{n-1}$ is $(-1)^k$ times the determinant of the submatrix obtained by deleting the kth column. This generalized cross product enjoys many of the familiar properties of the cross product in 3-space: It is a multilinear, skew symmetric function. The norm of the product, $v_1 \times v_2 \times \cdots \times v_{n-1}$, is the $(n-1)$-dimensional volume (or measure) of the parallelopiped spanned by the vectors $v_1, \ldots, v_{n-1}$. The vector $v_1 \times v_2 \times \cdots \times v_{n-1}$ is perpendicular to v_k for $k = 1, 2, \ldots, n-1$.

This raises the following question: given integers k and n such that $2 < k < n-1$, can we define a cross product of any k-tuple of vectors $v_1, v_2, \ldots, v_k$ in R^n having similar properties? To make the question precise, let us demand the following properties, similar to those in Theorem II:

(a) $v_1 \times v_2 \times \cdots \times v_k$ is a continuous function of the ordered k-tuple $(v_1, \ldots, v_k)$.
(b) $(v_1 \times v_2 \times \cdots \times v_k) \cdot v_i = 0$ for $i = 1, 2, \ldots, k$.
(c) If the vectors $v_1, v_2, \ldots, v_k$ are linearly independent, then $v_1 \times v_2 \times \cdots \times v_k \neq 0$.

The answer, strangely enough, is that such a cross product does not exist, with a single exception: $n = 8$ and $k = 3$. For the proof of this, the reader is referred to a paper by George Whitehead [6]; for explicit formulas for a cross product in this case, see Zvengrowski, [7].

Another property of the cross product of vectors in 3-space is the following: For any rotation r (i.e., orthogonal transformation of determinant $+1$) of 3-space and vectors v and w,

$$r(v \times w) = (rv) \times (rw). \tag{7}$$

(Note that this equation is *not* true if v is an orthogonal transformation of determinant -1). One can now prove the following:

PROPOSITION. *Assume that $n > 2$ and a cross product product is defined in R^n which is bilinear and satisfies equation (7) for any rotation of R^n; then $n = 3$.*

The proof depends on a knowledge of the real representations of the special orthogonal group $SO(n)$; we do not have space in this note to go into details.

This preservation of the cross product by rotations, expressed by equation (7), is less well known than the usual properties which are treated in our first theorem. It is normally only treated in advanced texts in theoretical physics or geometry.

References

1. J. F. Adams, On the nonexistence of elements of Hopf invariant one, Ann. of Math., 72 (1960) 20–104.
2. E. Calabi, Construction and properties of some 6-dimensional almost complex manifolds, Trans. Amer. Math. Soc. 87 (1958) 407–438.
3. B. Eckmann, Stetige Losungen linearer Gleichungssysteme, Comment. Math. Helv., 15 (1942/43) 318–339.
4. A. Hurwitz, Uber die Komposition der Quadratischer Formen von beliebig vielen Variablen, Nachr. Ges. d. Wiss. Göttingen, 1898, 309–316 (Math. Werke, Bd. II, 565–571).
5. N. Jacobson, Basic Algebra I, Freeman, San Francisco, 1974.
6. G. W. Whitehead, Note on cross sections in Stiefel manifolds, Comment. Math. Helv., 37 (1962/63) 239–240.
7. P. Zvengrowski, A 3-fold vector product in R^8, Comment. Math. Helv., 40 (1965/66) 149–152.
8. P. Hilton, General Cohomology Theory and K-theory, Cambridge University Press, Cambridge, 1971 (London Mathematical Society Lecture Note Series 1).

A "COUNTEREXAMPLE" FOR THE SCHWARZ-CHRISTOFFEL TRANSFORM

ELGIN JOHNSTON
Department of Mathematics, Iowa State University, Ames, IA 50011

The Schwarz-Christoffel Transform is a formula for a one-to-one analytic function that maps the upper half of the complex plane onto the inside of a polygon. As such the Schwarz-Christoffel Transform can be used to translate problems set in polygonal domains to more manageable problems set in the upper half plane. Such applications of the Schwarz-Christoffel Transform are common in problems involving two-dimensional flows, diffusion, potentials, etc. (see **[1]**, [3], **[5]**).

A statement of the Schwarz-Christoffel Transform can be found in most text books on elementary complex variables. One such statement is [4, p. 178]: *The functions $w = F(z)$ that map the upper half plane conformally onto polygons with interior angles $\pi\alpha_k$ $(k = 1,2,\dots, n)$ are of the form*

$$(1)\qquad F(z) = A\int_0^z \prod_{k=1}^{n} (\mathfrak{S} - x_k)^{-\beta_k}\, d\mathfrak{S} + B$$

where $x_1 < x_2 < \cdots < x_n$ are points on the real axis, $\beta_k = 1 - \alpha_k$ $(k = 1,2,\dots, n)$ and A, B are complex constants. Since the sum of the exterior angles of a polygon is 2π, we have:

$$(2)\qquad \beta_1 + \beta_2 + \cdots + \beta_n = 2 \quad \textit{with } -1 \leqslant \beta_k \leqslant 1.$$

(The polygons with one or more $\beta_k = \pm 1$ are those with some vertices at ∞ and/or some interior angles of 2π).

It is well known that conditions (1) and (2) are necessary for $F(z)$ to map the upper half plane conformally onto a polygon. However it is rarely stressed that these conditions are not sufficient for such a mapping and there appear to be few (if any) examples to this effect. In this paper we produce an example of a function $F(z)$ that satisfies (1) and (2) but is not one-to-one in the upper half plane. We start with a third condition (involving the choice of the x_k's) that is necessary for $F(z)$ to be univalent.

The following theorem gives some guidelines for the choice of the real numbers x_k mentioned in the statement of the Schwarz-Christoffel Transform.

THEOREM. *Let $\{\beta_k\}_{k=1}^n$ be given with $-1 \leqslant \beta_k \leqslant 1$ and let $x_1 < x_2 < \cdots < x_n$ be real. If the function $F(z)$ defined by* (1) *is univalent in the upper half plane, then*

$$(3)\qquad \left|\sum_{k=1}^{n} \frac{\beta_k}{z_0 - x_k}\right| \leqslant \frac{3}{\operatorname{Im} z_0}$$

whenever $\operatorname{Im} z_0 > 0$.

The theorem is an immediate consequence of the following lemma which gives a necessary

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P*, L*. Problem Solving in the Mathematics Curriculum: A Report, Recommendations, and an Annotated Bibliography. Alan H. Schoenfeld. MAA, 1983, ii + 137 pp, $5 (P). Report based on responses to the 1981 "Survey of Problem Solving Courses" conducted by the MAA-CUPM Subcommittee on Problem Solving. The Subcommittee writes: "...We believe that the primary responsibility of mathematics faculty is to teach their students to think: to question and to probe, to get to the mathematical heart of the matter, to be able to employ ideas rather than simply to regurgitate them." They recommend "that a series of problem solving courses at various levels of sophistication be developed and made regular offerings in the standard curriculum." This volume, complete with suggestions and bibliography, will serve as a valuable resource for moving in that direction. LCL

General. The Thread: A Mathematical Yarn. Philip J. Davis. Birkhauser Boston, 1983, 126 pp, $12.95. [ISBN: 3-7643-3097-X] An enthralling albeit mostly non-mathematical series of digressions on the origins of the name of P.L. Tschebyscheff, strung together from esoteric facts, vigorous opinion, and personal anecdotes covering an astonishing range of historical, linguistic, religious, and--occasionally--mathematical details. LAS

Precalculus, T(13: 1). Beginning Algebra, Third Edition. Alfonse Gobran. Prindle, Weber & Schmidt, 1983, viii + 391 pp. [ISBN: 0-87150-349-2] This edition retains a strong emphasis on solving word problems. The few revisions include additional examples and exercises. (TR, First Edition, May 1975.) JNC

Precalculus, T(13: 1). Algebra and Trigonometry for College Students, Second Edition. Richard S. Paul, Ernest F. Haeussler, Jr. Reston Pub, 1983, xii + 564 pp. [ISBN: 0-8359-0178-5] Uses simple vocabulary and a conversational style; changes include a rewriting of the chapter on real numbers, presentation of inequalities before functions and the addition of complex numbers, determinants, nonlinear systems of equations, linear equations in two variables, exponential and logarithmic functions, and sequences and series. (TR, First Edition, October 1978.) JNC

Foundations, S, P, L. The Nature of Mathematical Knowledge. Philip Kitcher. Oxford U Pr, 1983, ix + 287 pp, $25. [ISBN: 0-19-503149-0] A new proposal for a philosophy of mathematics, rejecting the tradition of mathematical truth as a priori. Kitcher argues that mathematics is not about abstract, Platonic objects but about "structures present in physical reality," that knowledge of individual mathematicians is explained by the knowledge passed on to them by authorities, and that this knowledge is warranted by a chain of authority and teaching ultimately grounded in primitive perceptual and empirical reality. As is common in philosophers' analyses of mathematics, supporting examples are drawn from a highly selective and well-known subset of mathematics (arithmetic, classical analysis, set theory, classical geometry), leaving the reader to wonder whether the theory is at all applicable to such things as high dimensional manifolds, forcing, nonstandard reals, or NP-completeness. LAS

Linear Algebra, T(15-17: 1, 2), L. Applied Matrix Algebra in the Statistical Sciences. Alexander Basilevsky. Elsevier North-Holland, 1983, xiii + 389 pp, $39.50. [ISBN: 0-444-00756-3] A systematic development of some of the more specialized theorems of linear algebra which are widely used in applications. Fairly complete proofs of the theorems are given. AO

Calculus, T(13: 2).** Single-Variable Calculus. Robert A. Adams. Addison-Wesley, 1983, xvii + 590 pp, $24.95. [ISBN: 0-201-10053-3] A clearly written, but refreshingly compact volume which thoroughly covers all the traditional topics; includes chapters on infinite series and power series but no multivariable topics. JNC

Calculus, S?. Averages: A New Approach. Jane Grossman, Michael Grossman, Robert Katz. Archimedes Foundation, 1983, vi + 61 pp, $3 (P). Monograph based on ideas in the authors' Non-Newtonian Calculus (TR, May 1973), and The First System of Weighted Differential and Integral Calculus (TR,

June-July 1981). Provides a general theory of averages of functions, but without any real justification for its use. RSK

Complex Analysis, P. Lecture Notes in Mathematics-971: Kleinian Groups and Related Topics. Ed: D.M. Gallo, R.M. Porter. Springer-Verlag, 1983, 117 pp, $8.50 (P). [ISBN: 0-387-11975-2] Proceedings of a workshop held at Oaxtepec, Mexico, August 10-14, 1981. JAS

Differential Equations, T(14), S, L. Introduction to Dynamics. Ian Percival, Derek Richards. Cambridge U Pr, 1982, ix + 228 pp, $34.50; $14.95 (P). [ISBN: 0-521-23680-0; 0-521-28149-0] A text that introduces the qualitative theory of differential equations and stability theory. Applications to biology and physics. Includes many exercises. JG

Differential Equations, P. Nonlinear Partial Differential Equations and their Applications: Collège de France Seminar. Ed: H. Brezis, J.L. Lions, D. Cioranescu. Pitman Pub, 1980. Volume I, 388 pp, $27.50 (P) [ISBN: 0-273-08491-7]; Volume II, 1982, 398 pp, $27.50 (P). [ISBN: 0-273-08541-7] These two volumes comprise written versions of lectures presented at the weekly seminar on applied mathematics at the Collège de France during the academic years 1978-1979 and 1979-1980. JAS

Numerical Analysis, T*(15-17: 1, 2). Computational Methods in Elementary Numerical Analysis. J. Ll. Morris. Wiley, 1983, xii + 410 pp, $41.95; $19.95 (P). [ISBN: 0-471-10419-1; 0-471-10420-5] Covers the topics traditionally included in an undergraduate course with an emphasis on the computational rather than the theoretical aspects of the subject. AO

Functional Analysis, P. Analytic Functional Calculus and Spectral Decompositions. Florian-Horia Vasilescu. Math. & Its Appl., V. 1. D Reidel Pub, 1982, xiv + 378 pp, $78.50. [ISBN: 90-277-1376-6] Spectral theory of finite commuting systems of operators on a Frechet space. The analytic functional calculus is an algebra homomorphism from the algebra of germs of analytic functions in neighborhoods of the joint spectrum of the system into a commutative algebra. Chapter I is background material, Chapter V is applications. Translation from the Rumanian. PZ

Analysis, S(17), P. Exterior Differential Systems and the Calculus of Variations. Phillip A. Griffiths. Progress in Math., V. 25. Birkhauser Boston, 1983, ix + 335 pp, $30. [ISBN: 3-7643-3103-8] A monograph based on lecture notes that utilizes techniques from the theory of exterior differential systems to shed new light on classical problems. A good foundation in general manifold theory is a prerequisite. JG

Geometry, T*(16: 1), S, P. Notes on Geometry. Elmer G. Rees. Universitext. Springer-Verlag, 1983, viii + 109 pp, $14 (P). [ISBN: 0-387-12053-X] Topics in Euclidean, projective and hyperbolic geometry are succinctly presented via Klein's Erlanger program definition of geometry; appropriate for a senior seminar; requires main ideas from linear algebra, group theory; metric spaces and complex analysis. Exercises. JNC

Optimization, T(16-18: 1, 2), S, L. Nonlinear Programming: Theory, Algorithms, and Applications. Garth P. McCormick. Wiley, 1983, xvii + 444 pp, $44.95. [ISBN: 0-471-09309-2] The bulk of the book presents the traditional topics that make up a first course in nonlinear programming as well as some of the newer techniques. Optimization models of several real world problems that can be solved using nonlinear programming are also given. AO

Optimization, T(14-16: 1), S, L. Linear Programming. Michel Sakarovitch, John B. Thomas. Springer-Verlag, 1983, xi + 206 pp, $19.80 (P). [ISBN: 0-387-90829-3] A relatively brief, self-contained, and mathematically rigorous introduction to linear programming. AO

Optimization, S(15-17), P. Linear Optimization and Approximation: An Introduction to the Theoretical Analysis and Numerical Treatment of Semi-infinite Programs. Klaus Glashoff, Sven-Åke Gustafson. Appl. Math. Sci., No. 45. Springer-Verlag, 1983, ix + 197 pp, $19.80 (P). [ISBN: 0-387-90857-9] The main focus of this book is the mathematical treatment of semi-infinite linear optimization problems (i.e., problems with a finite number of variables but an arbitrary number of linear inequality constraints). Computational techniques for problems of this type are presented. AO

Probability, P. Lecture Notes in Mathematics-972: Nonlinear Filtering and Stochastic Control. Ed: S.K. Mitter, A. Moro. Springer-Verlag, 1982, viii + 292 pp, $17 (P). [ISBN: 0-387-11976-0] Proceedings of the third 1981 session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held at Cortona, July 1-10, 1981. JAS

Statistics, P*. Computer Science and Statistics: Proceedings of the 14th Symposium on the Interface. Ed: Karl W. Heiner, Richard S. Sacher, John W. Wilkinson. Springer-Verlag, 1983, xi + 313 pp, $22 (P). [ISBN: 0-387-90835-8] Papers presented at a symposium held at Rensselaer Polytechnic Institute in July, 1982. Contains the keynote address by John Tukey, papers from thirteen workshop sessions, and fourteen more contributed papers. RSK

Statistics, T(15-17: 1), S*, P, L*. Understanding Robust and Exploratory Data Analysis. Ed: David C. Hoaglin, Frederick Mosteller, John W. Tukey. Wiley, 1983, xvi + 447 pp, $37.95. [ISBN: 0-471-09777-2] In the Wiley Series in Probability and Mathematical Statistics. Emphasizes the rationale and development of the techniques of exploratory data analysis and of robust and resistant techniques. Illustrates their use on generally small but real data sets. Contains exercises and good sets of references. RSK

Statistics, P*. Classification, Pattern Recognition, and Reduction of Dimensionality. Ed: P.R. Krishnaiah, L.N. Kanal. Handbook of Statistics, V. 2. Elsevier North-Holland, 1982, xxii + 903 pp, $108.75. [ISBN: 0-444-86217-X] Second volume in this series of reference books in statistical methodology and applications (Series and Volume 1, Analysis of Variance, TR, June-July 1981). Contains forty-one chapters on "discriminant analysis, clustering techniques and software, multidimensional scaling, statistical, linguistic and artificial intelligence models and methods for pattern recognition and some of their applications, the selection of subsets of variables for allocation and discrimination, and reviews of some paradoxes and open questions in the areas of variable selection, dimensionality, sample size and error estimation." RSK

Statistics, S(13-15), L. How to Tell the Liars from the Statisticians. Robert Hooke. Popular Statistics, V. 1. Dekker, 1983, xv + 173 pp, $15.95. [ISBN: 0-8247-1817-8] First volume in a new series on statistics for non-mathematicians. A collection of short anecdotes pointing out flaws in the presentation of statistical ideas in a wide variety of areas in everyday life. Somewhat reminiscent of Darrell Huff's How to Lie with Statistics. RSK

Computer Programming, S?(13-14). Graphic Software for Microcomputers. B.J. Korites. Kern Pub, 1981, 184 pp, $21.95 (P). [ISBN: 0-940254-01-8] A presentation of a coordinated series of Apple BASIC programs for a two- and three-dimensional interactive graphics package. Elementary vector and matrix explanations are given but not related to more general geometric theory. Some practice problems are provided as well as some discussion of key issues in translating to other computers and languages. No index; printing occasionally fails to reproduce small parts of dot-matrix printouts. JAS

Computer Programming, T(13-18: 1), S. PASCAL Programs for Scientists and Engineers. Alan R. Miller. Sybex, 1981, xxi + 374 pp, $16.95 (P). [ISBN: 0-89588-058-X] Contains many programs which implement standard, useful numerical algorithms. Chapters on mean and standard deviation, vector and matrix operations, simultaneous linear equations, curve-fitting, sorting, Newton's method, numerical integration. Chapter summaries. One appendix on reserved words and functions and another containing a summary of Pascal. Bibliography. Index. RJA

Computer Programming, T(13-18: 1), S. Structured Fortran 77 for Engineers and Scientists. D.M. Etter. Benjamin/Cummings Pub, 1983, xv + 357 pp, (P). [ISBN: 0-8053-2520-4] Chapter formats are very attractive. Each begins with the statement of an interesting, even unusual for a beginning FORTRAN text, applied problem whose programmed solution is given within the chapter. Each chapter contains summaries, key words and debugging aids sections, style and technique guides, sets of problems. Two appendices: summmary of FORTRAN 77 statements; and a list of FORTRAN 77 intrinsic functions. Glossary of key words. Answers to selected problems. Index. RJA

Computer Programming, T(13-18: 1), S. Programming in Modula-2, Second, Corrected Edition. Niklaus Wirth. Texts and Mono. in Comp. Sci. Springer-Verlag, 1983, 176 pp, $13.95. [ISBN: 0-387-12206-0] Provides an introduction to programming in general as well as a manual for learning Modula-2. Two appendices and an index (First Edition, TR, June-July 1983). RJA

Computer Science, S(16-18), P, L. Computing Skills and the User Interface. Ed: M.J. Coombs, J.L. Alty. Computers & People Ser., No. 3. Academic Pr, 1981, xii + 499 pp, $46. [ISBN: 0-12-186520-7] Addresses problems associated with communication between man and computer. Divided into three parts: (1) needs of computer users; (2) nature and acquisition of computing skills; (3) design of the user interface. Index. RJA

Computer Science, P. Lecture Notes in Computer Science-147: RIMS Symposia on Software Science and Engineering. Ed: Eiichi Gotto, et al. Springer-Verlag, 1983, 232 pp, $11.50 (P). [ISBN: 0-387-11980-9] This volume contains 12 research papers presented at the annual computer science symposia held in Kyoto, Japan during the period 1980-1982. The papers span a wide range of computer research including programming language design, analysis of algorithms, database systems, and computer architecture. MS

Applications (Physics), P. Theory of Phase Transitions: Rigorous Results. Ya. G. Sinai. Intern. Ser. in Natural Philo. Pergamon Pr, 1982, viii + 153 pp, $27. [ISBN: 0-08-026469-7] Developed from a series of lectures, this monograph surveys many of the rigorous results about the theory of phase transitions. The intended audience is "specialists in statistical physics." AO

Reviewers

RJA: Richard J. Allen, St. Olaf; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

The Mathematical Association of America
The Sixty Third Summer Meeting of the Association
State University of New York, Center at Albany

The Sixty Third Summer Meeting was held on the campus of the State University of New York, Albany Center during the period August 7-12, 1983. There were 909 registrants including 594 members of the Association. The Program Committee consisted of C. E. Burgess, Chairman; Donald R. Cohen, Vincent F. Cowling, Ronald L. Graham, William F. Hammond, and Theodore Bick. The talks for which abstracts were submitted consisted of the following:

Some Connections Between Algebra and Set Theory, by Barbara Osofsky, Rutgers University.

The talk began with the standard axioms of Zermelo-Fraenkel set theory (ZF). The axiom of choice (AC) was added as "Every subspace of a vector space has a complimentary subspace." Other algebraic statements closely related to AC were discussed. A module M over R[X,Y,Z] was constructed with the continuum hypothesis (CH) equivalent to $2^{\aleph_n} = \aleph_{n+1}$ for n contained in w were given. The existence of measurable cardinals was presented in terms of a mapping property of abelian groups. Lastly the proposition: "If every exact sequence of abelian groups 0 -> Z -> E -> G -> 0 splits, then G is free," which is undecidable in ZF + AC + CH was discussed.

Understanding and Teaching Mathematical Problem Solving, by Alan Schoenfeld, University of Rochester.

In a recent report on "Problem Solving in the Mathematics Curriculum" (MAA Notes #1), the MAA Committee on the Teaching of Undergraduate Mathematics wrote: "We endorse any approach to teaching mathematics that...actively engages students in the process of doing mathematics. We encourage the use of a "problem based approach" wherever possible in standard course offerings, including the participation of students in discussing, solving, and presenting their solutions to problems... . We recommend that a series of problem solving courses at various levels of sophistication be developed and made regular offerings in the curriculum." This talk focused on the results of research into mathematical problem solving (what accounts for the success of good problem solvers, the failures of poor ones?) on some practical suggestions for offering problem solving classes, and on some useful suggestions for implementing CTUM's recommendations.

Have You Ever Met a Polyhedron You Did Not Like?, by Branko Grunbaum, University of Washington.

Ancient Greeks started the study of polyhedra, but the topic remains rather confused to this day. There is no precise and agreed definition as to what is a polyhedron, and different authors have in mind different things when they speak of polyhedra. The practice of concentrating on regular or other very special polyhedra adds to the confusion: for them the need for workable and precise definitions is lessened, since many different approaches lead to quite similar results. Several concepts of "polyhedra" were discussed. This provided a convenient setting for presenting various old and new results, and many unsolved questions.

Theoretical and Experimental Contributions to the Development of a Science of Vehicular Traffic, by Robert Herman, University of Texas.

A brief historical survey was given to show that traffic problems are very old. In addition, a bird's eye view of developments in the science of vehicular traffic will touch on single lane flow and stability, a kinetic theory of multilane flow, the characterization of urban traffic and energy consumption, and a two-fluid model of town traffic.

Progress Report of the Commission on Precollege Education in Mathematics, Science, and Technology, by Katherine P. Layton, Beverly Hills High School.

In response to the current decline in the quality and quantity of precollege mathematics and science education in the United States, the National Science Board established the Commission on Precollege Education in Mathematics, Science and Technology. The purpose of the Commission is to define a national agenda for improving mathematics and science education. A progress report was given explaining the work done by the Commission during its first fourteen months. Included was a discussion of tentative recommendations.

Convexity Ideas in Geometric Function Theory, by Thomas H. MacGregor, SUNY, Center at Albany.

Presented were ways in which the idea of convexity plays a role in complex analysis. In particular, the use of linear methods in the area of conformal mapping was discussed. The extreme points and the convex hulls of various families of analytic functions relate these sets to the solution of extremal problems and to the set of support points were described. This involved a blend of geometric, algebraic and analytic notions.

Primality Testing, by Carl Pomerance, University of Georgia.

If a large number is prime, how can one quickly find a proof that it is so? With recent tests, average 100 digit primes take about 33 seconds on a fast computer, while 200 digit primes take about 8 minutes. Worst cases take about twice as long. With the older tests, worst case 100 digit primes could have taken about 100 years of computer time and 200 digit primes about 10^9 years.

Nevertheless, much of the new test has its roots in the older tests. These developments were discussed, especially the similarities and differences between the new tests and the old tests.

Galois' Version of Galois Theory, by Harold M. Edwards, NYU--Courant Institute.

This talk was an exposition of the ideas of Galois' original memoir "On the conditions for the solvability of equations by radicals," including his method of defining the Galois group of an equation and his proof that an equation is solvable by radicals if and only if its group is solvable.

Other speakers and their titles were: Richard Lewontin, Harvard University, "Analysis of Complex Generic Systems," and Elias M. Stein, Princeton University, "Some Ideas in the Development of Fourier Analysis," the Earle Raymond Hedrick Lecture for 1983.

Special Sessions

Mini-Courses: The Association sponsored mini-courses entitled "Pascal for Mathematicians," by Harley Flanders, Florida Atlantic University; "Problems from Industry for Use in the Undergraduate Classroom," by Jeanne L. Agnew and Marvin S. Keener, Oklahoma State University; "An Introduction to the Mathematical Techniques and Applications of Computer Graphics," by Joan Wyzkoski, Bradley; "Conduit Microcomputer Software in Mathematics Instruction," by David A. Smith, Duke University; "Commercial Microcomputer Software in Undergraduate Mathematics Instruction," by David A. Smith, Duke University; "Coloring Problems," by David M. Berman, University of New Orleans.

Contributed Paper Sessions: The Association sponsored three contributed papers sessions. The interest areas and the list of presentors and submitted abstracts follows:

The Use of Computers in Undergraduate Mathematics Instruction. Presider: Ronald H. Wenger, University of Delaware.

"How Effective is Computer-Assisted Instruction for Mathematics Remediation," by Henry Africk, New York City Technical College of the City University of New York.

Computer-assisted instruction can improve overall student performance when used as a supplement to regular classroom instruction. The amount of improvement will vary depending on the topic, the type of student, the quality of classroom instruction as well as the instructional design of the computer program itself. These points were illustrated with examples from the computer-assisted instruction project at New York City Technical College.

"The Use of a Software Authoring System to Individualize Instruction in a Mathematics Pre-Calculus Course," by Fredric Tufte, University of Wisconsin, Platteville, Wisconsin.

This paper discussed the use of TIPS (Teaching Information Process System), a software authoring system, in the design of a computer-assisted instructional package to supplement normal instruction in a mathematics pre-calculus course. After being designed and implemented by the instructor, the student is able to take multiple choice tests over small units of material. The system diagnoses concepts with which the student has difficulty and prescribes additional work for the student to hopefully correct the exhibited deficiencies.

"A Computer Aided Calculus Course," by Mario Martelli, Bryn Mawr College, Bryn Mawr, Pennsylvania.

Described was how the Computer-Calculus course organized in the last two years. A set of programs designed for addressing the main topics of a first year Calculus, which require a minimum knowledge of computer language, and which can be used both for illustration or investigation purposes was also described.

"Motivating Trigonometry: Using the Computer to Investigate Harmonic Motion," by Martin E. Flashman, Humboldt State University, Arcata, California.

The harmonic motion of a spring can motivate the circular trigonometric functions using the standard second order differential equation $y'' = -y$ with $y(0) = 0$ and $y'(0) = 1$. Proceeding by Euler's Method, computations illustrate delayed feedback and the important interaction of y and y'. With Δt small, a computer or hand plot of the approximate solutions for $y(t)$ and $y'(t)$, with $x(t) = y'(t)$, suggests the relations of the points $(x(t),y(t))$ to the unit circle, which then can be proven. Finally the parameter t can be discovered as the arc length on the unit circle.

"Using a Surface Plotting Program in Sophomore Calculus," by Richard J. Cleary, Saint Michael's College, Winooski, Vermont.

In the spring of 1983 at Saint Michael's College, fourth semester calculus students were required to complete a computer graphics project involving the use of a surface plotting program. The program, called BDPLOT, was developed by the National Center for Atmospheric Research. Our version was adapted by Middlebury College and run on Saint Michael's DEC PDP-11 using the RSTS/E operating system. The primary goal of the project was to allow the students to obtain a graph of function $f(x,y)$ which they had previously analyzed to find critical points. Each student submitted a copy of three pictures with the analysis of the function attached. The results of the project and

suggestions for improvement are discussed.

"A Numerical Approximation Method for Double Integrals," by Donald R. Snow, Brigham Young University, Provo, Utah.

This numerical method arose in a calculator/computer multivariable calculus class. It is the natural generalization of the Trapezoidal Rule for one-variable integrals and was worked out through a series of homework problems and discussions with the class. The domain of integration for the double integral is divided into triangles and the function surface is approximated by a plane through the function values at the corners of these triangles. The "Triangular Rule" is then the sum of the volumes of each of these prisms. The formulas turn out to be quite simple and give good results. By experimenting with the Romberg correction formulas the error estimate for the Rule was determined. This is a good illustration that through the use of computational aids in the class the students can understand more clearly the concept under consideration and can share in the excitement of discovery.

"Inverse Functions, Implicit Functions and Computer Graphics," by Sheldon P. Gordon, Suffolk Community College, Selden, New York.

The use of sophisticated computer graphics provides the user with a powerful tool to investigate many aspects of mathematics that are inherently geometric in nature and yet cannot be dealt with by hand due to their complexity. The talk was focused on two such topics from elementary calculus and differential equations and demonstrated how the computer can add a valuable dimension to student understanding.

"The St. Petersburg Paradox," by Allan J. Ceasar, U.S. Merchant Marine Academy, Kings Point, New York.

The Strong and Weak Law of Large Numbers was compared in this paper through the use of the St. Petersburg Paradox. The gamble, which was shown to have an infinite mean, created controversy because most participants would be willing to pay only a relatively small entrance fee to play. This is inconsistent with the concept of a fair game. Utility and expected utility was discussed as a means of avoiding this dilemma. A computer program to simulate this gamble was analyzed and a graphical presentation of the computer printout proved to be a valuable pedagogical tool in demonstrating the Law of Large Numbers.

"Using Computer Simulation in Teaching Statistics," by Robin H. Lock, St. Lawrence University, Canton, New York.

The computer may be effectively used in most statistics courses as a tool for generating random data. It allows the instructor to control the experimental situation completely by specifying both the "true" underlying model and the nature of the random noise which was introduced into the sampling scheme. Furthermore, the student may easily obtain samples, even for a complicated experimental design, in a convenient form. Two examples of this sort of computer simulation were discussed. The first is a rather simple program called SAMPLER which generates sample from a randomly chosen distribution. The student's task is to identify the distribution and estimate its parameters. The second program, GOLFBALLS, simulated both the distance and deviation of golf shots. The student may control variables such as the color, compression, and construction of the balls to investigate their effect on the shots.

"An Undergraduate Course in Time Series Analysis Made Possible with Minitab," by Constance M. Elson, Ithaca College, Ithaca, New York.

Formal time series analysis involves graduate-level mathematics by using Minitab to provide "hands-on experience," it proved possible to offer a one-semester course in the subject to mathematically unsophisticated students. Typically, these students had had 1-2 semesters of calculus and 1 semester non-calculus statistics course. Minitab, a widely-used software package for statistical analysis of data, is quickly mastered and easy to use in an interactive mode; it proved valuable in teaching some basic concepts of stochastic modeling which would not have been accessible to these students if stated in abstract terms. By using Minitab to do their own "random sampling" of a random variable with a given probability distribution, the difference between complete theoretical knowledge and inferences based on partial observations became explicit and vivid. Building on this, it was possible to convey a correct intuitive understanding of the difference between a stochastic process (a sequence of random variables) and a single realization of the process, the "observed time series" (a single sample path of the process). Once the students had grasped these ideas, concepts like stationarity, ergodicity, and conditional expectation became meaningful to them.

<u>Classroom Notes</u>. Presider: J. Arthur Seebach, St. Olaf College.

"Using Discovery and Invention to Learn About Computers," by Marjory Baruch, Hamilton College, Clinton, New York.

"Simpson's Rule as a Weighted Average," by Tom Tucker, Colgate University, Hamilton, New York.

"An Heretical Approach to Solving Polynomial Equations," by Sylvan Burgstahler, University of

Minnesota, Duluth, Minnesota.

"The Predator-Prey Problem Revisited," by Roland Lamberson, Humboldt State University, Arcata, California.

"Numerical Linear Algebra: A Brief Survey," by Steven Leon, Southeastern Massachusetts University.

"An Elementary Proof That Certain Numbers are Irrational," by Harvey Hindin, Hi-Tech Editorial, Inc., Dix Hills, New York.

The Undergraduate Mathematics Curriculum. Presider: Jerome A. Goldstein, Tulane University.

"A Multiple Linear Regression Model for Placement Using the MAA Placement Exam," by E.J. Manfred and C.F. McCarthy, U.S. Coast Guard Academy.

This was a continuation of a study designed to investigate methods by which the use of available data on incoming cadets at the United States Coast Guard Academy can be used as a guide for placement in the first year mathematics course. A multiple linear regression model that predicts grades in the first year mathematics course was used for the past several years. A new model was developed and included as one of the independent variables the score of the MAA Calculus Readiness Placement Exam. Intercorrelations between Calculus Readiness Exam and other independent predictor variables were examined.

"Mathematics Remediation with Dignity," by L. Clark Lay, Aurora College, Aurora, Illinois.

Optimum college mathematics remediation requires more than just a review of concepts met before but not mastered. Fresh content is essential. This must be new, simple, and significant.

"The Undergraduate Teaching Assistant in Development and College Mathematics Courses," by Marguerite Graves, Pennsylvania State University, Allentown, Pennsylvania.

The peer tutor, i.e., PT, the undergraduate teaching assistant fairly close to the student in academic development, is one of our precious resources often preferable to the graduate TA even when available.

In various forms of individualized instruction or tutorial set-ups, the PT has a well defined function, grading unit tests, answering questions, and facilitates flexibility in the curriculum, etc. In the "group" method, the role is more subtle. A structure is created when the better student helps the members of the group--and thus him or herself. Various group projects might include take home tests and grading of take home tests for the other groups, chapter outlines, etc.

"Variables in the Description of Developmental Mathematics Programs," by Miriam Hecht, Hunter College, CUNY, New York.

Over the past decade, developmental mathematics programs have proliferated on American campuses. These programs, many of which have been described in the literature, often embody important educational innovations. However, the reader may be at loss to evaluate, replicate, compare, or even understand such programs because authors vary in their selection and discussion of salient variables. Thus, one account may describe the role of tutors and computers while failing to mention the size of the program, the content, or the completion rate. Another may provide details of an independent study program without considering the number, age, and preparation of the students or the vehicle of instruction.

In this paper, the author identified and described a number of variables entering into the design and implementation of developmental programs. Such an outline provided a framework for description and analysis. It is particularly useful in connection with developmental courses because these courses usually address the needs of nontraditional students and incorporate an impressive variety of resources. However, it is applicable to the description of any college program, in whatever field.

"Large Section Teaching of Developmental Mathematics," by John L. Tilley, Mississippi State University, Mississippi State.

The fiscal, logistical, and philosophical reasons for using our format will be touched on. As a mandated program by the Board of Trustees of the Institutions of Higher Learning for the State of Mississippi, the "tracking" of all students meeting the Board's cut-off requirements has produced a great amount of statistical data about these students. From this mass of data, information about students' grades in the course and subsequent grades in credit mathematics courses is available. Focus on subsequent courses is primarily on College Algebra with a semester prior to the introduction of Developmental Mathematics.

"Convergence Isn't Sufficient," by Joanne Growney, Bloomsburg State University, Bloomsburg, Pennsylvania.

Mathematics instruction primarily focuses on teaching students to find solutions to carefully posed problems. This process of "convergent thinking" is a necessary and effective thinking in the large number of day-to-day situations that are permeated with unidentified problems or questionable solution proposals. Such situations need individuals skilled in divergent thinking (the ability to produce a variety of ideas that lead in different directions) and evaluation (the ability to reach decisions about the quality of proposed solutions to a problem). Traditionally, mathematics courses have been seen as a "natural setting" for teaching thinking. However, learning to think is not an automatic by-product of learning mathematics; moreover, divergent thinking and evaluation are not automatic by-products of convergent thinking skills. Instead, it is likely that we teach thinking skills only when it is our goal to do so. Thus, it is important that we design our teaching strategies and curricular materials to develop specific thinking skills. Few teaching materials that emphasize divergent thinking or evaluation were found. The presentation provided examples of such materials and will urge interested mathematicians and teachers to develop instructional materials to help to achieve these teaching goals.

"A Mathematics--Art Course," by Frederick Solomon, SUNY College at Purchase, Purchase, New York.

The course taught at State University of New York, College at Purchase has both an artist and a mathematician as instructors. The course is aimed at freshmen and sophomores and combines mathematical ideas with visual representation. Students are required to complete ten projects visually illustrating various mathematical concepts. The talk included syllabus, annotated bibliography and, in general, focused on techniques for topic presentation and rationale for inclusion of topics in the course.

"The Sloan Foundation Discrete Mathematics Initiative," by Steve Maurer, Swarthmore College, Swarthmore, Pennsylvania.

The Sloan Foundation recently announced grants of $40,000 each to 6 institutions to develop mainstream mathematics curricula for the first two undergraduate years in which discrete mathematics receives as much attention as calculus. What programs have these 6 institutions proposed? Should this initiative succeed? Will it?

"Constraints on the Changing Curriculum," by Richard Anderson, Louisiana State University, Baton Rouge, Louisiana.

"Mathematics in China--Summer, 1983," by Donald Bushaw, Washington State University.

The Association also presented the following Film Program:

Points of view: Perspective and projection
Dragon fold...and other ways to fill space
Caroms
Linear programming
Turning a sphere inside out
Hypothesis testing, inferential statistics, Part II
Journey to the center of a triangle
Symmetry and tesselations

Board and Business Meeting

The Board of Governors met at 9:00 a.m. on Sunday, August 7, 1983. The major items of business transacted by mail will be announced to the membership in FOCUS.

The Business Meeting of the Association was held at 4:30 p.m. on Tuesday, August 9, 1983 and the Carl B. Allendoerfer and George Polya Awards for expository articles in Mathematics Magazine and the Two-Year College Mathematics Journal were presented.

Election of Members

At its meeting on August 7, 1983, the Board elected to membership 794 applicants for individual membership and 10 applicants for academic membership. The latter follows:

Juniata College, Huntingdon, Pennsylvania
Norwalk State Technical College, Norwalk, Connecticut
Pace University, Pleasantville, New York
Chattahoochee Valley Community College, Phenix City, Alabama
Ryerson Polytechnical Institute, Toronto, Ontario, Canada
University of Puerto Rico, Mayaguez, Puerto Rico
University of Southern Colorado, Pueblo, Colorado
Delaware County Community College, Media, Pennsylvania
Mercy College of Detroit, Detroit, Michigan
University of West Indies, Kingston, West Indies

Respectfully submitted,

David P. Roselle, Secretary

CENTER SECTION INDEX

Volume 90, 1983

Official Reports

[6, pg. 145] that for such $g(w)$, $\left|\frac{g'''(0)}{g'(0)}\right| \leqslant 18$. Based on this inequality, one could derive conditions concerning the first three derivatives of F analogous to the condition in the Theorem. Using such a result, one might be able to construct an example showing that (1), (2) and (3) are not enough to guarantee the univalence of F.

References

1. R. Churchill, Complex Variables and Applications, McGraw-Hill, New York, 1960.
2. E. Hille, Analytic Function Theory, vol. 2, Ginn, Boston, 1962.
3. J. E. Marsden, Basic Complex Analysis, W. H. Freeman, San Francisco, 1973.
4. Z. Nehari, Complex Analysis, Allyn and Bacon, Boston, 1967.
5. L. Pennisi, Elements of Complex Variables, Holt, Rinehart and Winston, New York, 1963.
6. Chr. Pommerenke, Univalent Functions, Vanderhoeck and Ruprecht, Göttingen, 1975.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

A CLASSROOM NOTE ON THE SAMPLE VARIANCE AND THE SECOND MOMENT

J. J. BIRCH AND TIM ROBERTSON

Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA 52242

In teaching statistics, one uses a number of fundamental criteria to select estimators and judge their quality. Apparent paradoxes occasionally arise when two or more of the principles are applied to the same problem. It is the purpose of this note to point out one such conflict that appears not to have been noticed by many statisticians.

The two criteria we have in mind are the principles of maximum likelihood and mean squared error. We consider the concepts in the selection of an estimator for the variance of a normal population. The normal distribution is a two parameter distribution and if we are interested in its variance, σ^2, then the mean, μ, is an auxiliary parameter. More specifically, the estimation of σ^2 depends upon whether or not μ is known; if μ is unknown, its estimation seems to take precedence over the estimation of σ^2. When μ is known, the maximum likelihood estimator of σ^2 is the average of the squared deviations from μ, i.e., the second sample moment about the mean

$$S^2(\mu) = \frac{1}{n}\sum (X_i - \mu)^2.$$

This estimator, when μ is known, has certain desirable properties such as: (1) it results from the method of moments; (2) it is unbiased and, moreover, (3) it is the minimum variance unbiased estimator.

The criterion of sufficiency is often used in choosing estimators. A statistic is sufficient if, roughly speaking, all pertinent sample information about the distribution is contained in the value of the statistic. Thus, if an estimator is a function of a sufficient statistic, then the full expression of the extent to which the data give information about the parameter is obtained from such an estimator. M. G. Kendall (1946) apparently had this criterion in mind when he stated in his classic text (page 11) that "A sufficient estimator is best for any sample size since it gives all the information about θ that a sample can give; and is most efficient for large samples." Continuing, he points out that when sampling from a normal population with known mean

$$S^2(\bar{X}) = \frac{1}{n}\sum(X_i - \bar{X})^2$$

is not sufficient for σ^2. Thus, he says "If we know the real mean μ, there is little point in preferring the sample variance $S^2(\bar{X})$ to the second moment $S^2(\mu)$ as an estimator of the parent variance."

Thus, $S^2(\mu)$ seems to be a very desirable estimator. However, before we can claim that this estimator is "better" than others, we should, to begin with, consider what losses will be sustained if errors are made. A popular method for comparing estimators, based on errors, is to compare their mean squared error (MSE).

The sample variance, $S^2(\bar{X})$, can be used as an estimator of σ^2 when μ is known although, as noted above, we might not expect it to compete very favorably with $S^2(\mu)$, the second sample moment, since it seemingly ignores our knowledge of μ. However, when we compare the MSE's of these two estimators, we find that

$$\text{MSE(of the sample variance)} = \left(\frac{2}{n} - \frac{1}{n^2}\right)\sigma^4 < \frac{2}{n}\sigma^4 = \text{MSE(of the second sample moment)}.$$

These values are obtained from the first two moments of the chi-squared distribution and the facts that $nS^2(\bar{X})/\sigma^2$ and $nS^2(\mu)/\sigma^2$ have chi-squared distributions with $n - 1$ and n degrees of freedom, respectively. This is the apparent paradox and it seems to imply that when μ is known, we should ignore this information and use the estimator, $\bar{X}$, for μ.

In connection with some recent research (Randles (1982) and Pierce (1982)), we note that the second sample moment, $S^2(\mu)$, is a U-statistic so that the sample variance $S^2(\bar{X})$ is a U-statistic with an estimated parameter. A natural question would be to ask how much efficiency is lost when using a U-statistic with an estimated parameter over the corresponding U-statistic. Of course the U-statistic is unbiased, by definition, but the corresponding U-statistic with an estimated parameter may be biased. Thus, efficiency cannot be measured in terms of variance. One way of measuring it would be in terms of mean squared error. This example illustrates that with regard to squared error loss the U-statistic with an estimated parameter may, in fact, have a smaller mean squared error than the corresponding U-statistic.

Another comment we make is that the mean squared errors of the sample variance and the second moment can be computed in general (cf. Cramér (1946)). It turns out that as long as the population kurtosis is at least as large as 2.5, the MSE of the sample variance is smaller than the MSE of the second sample moment. One might infer from this fact that the sample variance is more "robust" than the second sample moment.

When μ is unknown in a normal population, we usually compare $S^2(\bar{X})$, the sample variance, with the "best" unbiased estimator of σ^2. In this case, it is well known that the unbiased estimator

$$\frac{n}{n-1}S^2(\bar{X}) = \frac{1}{n-1}\sum(X_i - \bar{X})^2$$

is the minimum variance unbiased estimator of σ^2 (it is the recommended estimator in many books on statistics). An easy calculation yields

$$\text{MSE}\left(\frac{n}{n-1}S^2(\bar{X})\right) = \frac{2}{n-1}\sigma^4$$

and

$$\text{MSE}\left(\frac{n}{n-1}S^2(\bar{X})\right) > \text{MSE}(S^2(\mu)),$$

which agrees with our intuitive principle. Nevertheless, we still feel a little shaky about using $S^2(\mu)$ as an estimator of σ^2 when μ is known, because $S^2(\bar{X})$ has an even smaller MSE than $S^2(\mu)$.

One explanation is that none of these estimators has the correct divisor if we are to use mean

squared error as a measure of quality. It is well known and easy to verify that

$$\mathrm{MSE}\left(\frac{n}{n+1}S^2(\bar{X})\right) \leqslant \mathrm{MSE}(anS^2(\bar{X}))$$

and

$$\mathrm{MSE}\left(\frac{n}{n+2}S^2(\mu)\right) \leqslant \mathrm{MSE}(anS^2(\mu))$$

for every $a > 0$. Thus if mean squared error is our criterion, we should use

$$\frac{n}{n+2}S^2(\mu) = \frac{1}{n+2}\sum(X_i - \mu)^2$$

when μ is known (an admissible estimator) and

$$\frac{n}{n+1}S^2(\bar{X}) = \frac{1}{n+1}\sum(X_i - \bar{X})^2$$

when μ is unknown (an inadmissible estimator (cf. Stein (1964)). Moreover, our intuition is then confirmed as

$$\mathrm{MSE}\left(\frac{n}{n+2}S^2(\mu)\right) = \frac{2}{n+2}\sigma^4 < \frac{2}{n+1}\sigma^4 = \mathrm{MSE}\left(\frac{n}{n+1}S^2(\bar{X})\right).$$

Thus the statistic using the known mean μ is better, in the sense of mean squared error, than the one ignoring this information. It should be noted, however, that these two estimators of σ^2 are almost never considered in basic textbooks on statistics.

On the other hand, one might not be so much interested in σ^2 as in σ. Thus we might compare the mean squared errors of $S(\bar{X})$ and $S(\mu)$ as estimators of σ. The distributions of these statistics are not so tractable as those of $S^2(\bar{X})$ and $S^2(\mu)$. However, we calculated their mean squared errors for sample sizes of $1, 2, \ldots, 30, 40, 50, \ldots, 100$, and in each of these instances the mean squared error of $S(\mu)$ was less than that of $S(\bar{X})$. Thus it seems that for the estimators $S(\bar{X})$ and $S(\mu)$ of the standard deviation σ our intuition holds while, for the estimators $S^2(\bar{X})$ and $S^2(\mu)$ of σ^2 it does not hold.

Finally, mean squared error is not the only possible criterion for measuring the quality of an estimator. Among other measures of the "closeness" of an estimator to its parameter is one proposed by Pitman. It can be shown that

$$P\left[(S^2(\bar{X}) - \sigma^2) > (S^2(\mu) - \sigma^2)\right] \geqslant \frac{1}{2}$$

so that, in the Pitman sense, $S^2(\mu)$ is "closer" to σ^2 than $S^2(\bar{X})$ is. Here again our intuition is upheld.

References

1. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

2. M. G. Kendall, The Advanced Theory of Statistics, C. Griffen & Co., 3rd ed., vol. 2, 1946.

3. D. A. Pierce, The asymptotic effect of substituting estimators for parameters in certain types of statistics, Ann. Statist., 10 (1982) 475–478.

4. R. H. Randles, On the asymptotic normality of statistics with estimated parameters, Ann. Statist., 10 (1982) 462–474.

5. C. Stein, Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Ann. Inst. Statist. Math., 16 (1964) 155–160.

ANSWER TO PHOTO ON PAGE 676

J.H.C. Whitehead.

PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, DAVID BORWEIN (ADVANCED PROBLEMS), H. M. W. EDGAR (ELEMENTARY PROBLEMS), AND D. H. MUGLER

*Send all **proposed** problems, typed and in duplicate if possible, to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by April 30, 1984. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 3025. *Proposed by S. W. Graham, University of Texas at Austin and D. Hensley, Texas A & M University.*

Let $\lambda(n)$ be the function on the positive integers determined by the conditions that $\lambda(p) = -1$ for all primes p and $\lambda(mn) = \lambda(m)\lambda(n)$ for all m and n. (This is commonly referred to as Liouville's lambda function.) Show that there exists a $C > 0$ such that the number of $n \leqslant x$ for which $\lambda(n) = \lambda(n+1)$ exceeds Cx, for x sufficiently large.

E 3026. *Proposed by H. Kestelman, University College, London.*

Let A be a skew-symmetric matrix of order $2n$ considered as a function of its $n(2n-1)$ independent upper off-diagonal elements; these are complex numbers ordered conventionally as $x_1, x_2, \ldots, x_N$. Show that there exist matrices F and G whose elements are rational functions of the x_r's with rational (real) coefficients, and which are such that $A = FG$ and $\det F = \det G$.

E 3027. *Proposed by Cran Minh Crung, Skogn, Norway.*

Evaluate in closed form $B(r, m, n) = \sum_{k \equiv r (\mathrm{mod}\ m)} \binom{n}{k}$.

E 3028. *Proposed by B. Tomaszewski, University of Wisconsin.*

Let $\{z_i\}$ be a sequence of complex numbers which converges to 0. Prove that there exists sequence $\varepsilon_i = \pm 1$ such that $\sum_i \varepsilon_i z_i$ converges.

E 3029. *Proposed by S. W. Golomb, University of Southern California.*

Let α and β be nonzero elements of a field F which satisfy $\alpha + \beta = 1$ and $\alpha^{-1} + \beta^{-1} = 1$. Prove that $\alpha = \beta^{-1}$ and that $\alpha^6 = \beta^6 = 1$.

E 3030. *Proposed by R. E. Shafer, Berkeley, CA.*

Let $a_0 \geqslant a_1 \geqslant a_2 \geqslant \cdots \geqslant a_{N-1} > 0$. The polynomial $P_N(z) = \Sigma\, a_i z^i$ has zeros $|\zeta| \geqslant 1$. If for one of these zeros $|\zeta| = 1$, then there is $n \geqslant 2$, $n|N$, such that

$$P_N(z) = \frac{z^n - 1}{z - 1} \sum_{i=0}^{\frac{N}{n}-1} a_{ni} z^{ni}.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Convergent Subsequences

E2919. [1981, 764] *Proposed by Elgin Johnston, Iowa State University, Ames.*

(a) Let $\{a_k\}_{k=1}^{\infty}$ be a sequence of real numbers. Suppose that, for each prime p, the sequence $\{a_{np}\}_{n=1}^{\infty}$ is increasing and bounded above. Let $\{b_k\}$ be the sequence obtained when those elements with prime subscripts are deleted from $\{a_k\}$. Show $\lim_{k\to\infty} b_k$ exists.

(b) Suppose in part (a) the phrase "is increasing and bounded above" is replaced by "converges." Does the result still hold?

Solution by J. Browkin, Warsaw University, Poland. (a) From the assumptions it follows in particular that the limit of a_{2n} exists; call it A. Moreover, if composite numbers r, s, $r > s$, have a common prime factor, then

(1) $$a_r \geqslant a_s.$$

Take a composite number m and let q be its least prime factor. Then $m = tq$ with $t \geqslant [\sqrt{m}]$. Applying (1) three times we obtain

$$a_{2m} \geqslant a_m = a_{tq} \geqslant a_{2t} \geqslant a_{2[\sqrt{m}]}.$$

From $\lim_{m\to\infty} a_{2m} = A$, $\lim_{m\to\infty} a_{2[\sqrt{m}]} = A$, we conclude that the sequence a_m with composite indices m also converges to A, i.e., $\lim_{k\to\infty} b_k = A$.

(b) No. Take $a_n = 1$ if n is the square of a prime; $a_n = 0$ otherwise. Evidently, for a prime p, all terms but one in the sequence $\{a_{np}\}_{n=1}^{\infty}$ are equal to 0. Thus it converges. In the sequence b_k infinitely many terms are equal to 0, and infinitely many to 1. Thus it diverges.

Also solved by U. Abel (Germany), L. Boxer, D. Cantone (Italy), D. Goldberg, T. Hermann (Hungary), V. Hernandez (Spain), D. Hill, J. Hook (student), W. Janous (Austria), M. Josephy (Costa Rica), C.-N. Lee (student), G. S. Lessells (Ireland), O. P. Lossers (Netherlands), D. A. Mattson, R. E. Megginson, F. B. Miles, R. Moller, A. M. Nadel (student), W. A. Newcomb, E. Posti (Finland), H. Prodinger (Austria), D. A. Rawsthorne, B. Richter, J. H. Riley, Jr., T. Šalat (Czechoslovakia), E. A. Schwenk, A. Seeger (Germany), H.-J. Seiffert (Germany), D. K. Skilton (Australia), A. Smuckler (Israel), R. J. Snelling, J. Suck (Germany), G. G. Thompson, University of South Alabama Problem Group, D. G. Weinmann, D. Wells, P.-Y. Wu, and the proposer.

The Discriminant of $x^n - \Sigma_1^n x^{n-i}$

E 2920 [1982, 63]. *Proposed by J. O. Shallit, University of California, Berkeley.*

Let $f(x) = x^n - x^{n-1} - x^{n-2} - \cdots - x - 1$, $n \geqslant 2$.

(a) Show that the discriminant of f is given by

$$\operatorname{disc}(f) = (-1)^{n(n+1)/2}\left[\frac{(n+1)^{n+1} - 2^{n+1}n^n}{(n-1)^2}\right].$$

(b) Show that the fraction in (a) is an integer directly.

Solution to part (a) *by D. C. Kurtz (Zomba, Malaŵi) and P.-Y. Wu (Hsinchu, Taiwan).* The

formula is the special case $a = -1$ of problem 840, *Problems in Higher Algebra*, by D. K. Faddeev and I. S. Sominskii (W. H. Freeman, 1965).

Solution to part (b) *by most of the solvers listed below*. Replacing n by $m+1$, we have to show that for $m \geqslant 2$, $\operatorname{disc} f \equiv 0 \bmod m^2$. This is a consequence of the binomial theorem.

M. Josephy (Costa Rica) proved the result by noting that $p(1) = p'(1) = 0$ when $p(n) = (n+1)^{n+1} - 2^{n+1}n^n$.

Also solved by D. Bode (Germany), J. Brillhart, L. Carlitz, S. Gagola, H. Kappus (Switzerland), C.-N. Lee, O. P. Lossers (Netherlands), P. Ramankutty (Mexico), A. T. Steele, and the proposer.

Natural Numbering of a Lower Triangular Array

E 2926 [1982, 130]. *Proposed by Jerrold W. Grossman, Oakland University.*

The cells in a lower triangular array are indexed by pairs of integers (i, j), where $1 \leqslant j \leqslant i \leqslant n$. There are two natural ways in which to order the cells:

ROW ORDER: $(i, j) < (i', j') \Leftrightarrow i < i'$ or $(i = i'$ and $j < j')$.

COLUMN ORDER: $(i, j) < (i', j') \Leftrightarrow j < j'$ or $(j = j'$ and $i < i')$.

Let $S(n)$ be the number of cells whose rank is the same in both systems.

(a) Evaluate $S(n)$.

(b) Find the values of n for which $S(n) = 4$.

(c) Find the values of n for which $S(n)$ is odd.

Solution by O. P. Lossers, Eindhoven Institute of Technology, Netherlands, and independently by the proposer and Donald G. Malm, Oakland University (jointly). We shall assume that $n \geqslant 2$.

(a) Clearly cell (i, j) has row rank

$$1 + 2 + \cdots + (i-1) + j = \binom{i}{2} + j$$

and column rank

$$n + (n-1) + \cdots + (n-j+2) + i - j + 1 = \tfrac{1}{2}(j-1)(2n-j) + i.$$

Hence $S(n)$ is the number of solutions of

$$(1) \qquad i(i-1) + 2j = (j-1)(2n-j) + 2i, \quad 1 \leqslant j \leqslant i \leqslant n.$$

Putting $x = 2i - 3$, $y = 2n - 2j - 1$, it follows that

$$(2) \quad S(n) = |\{(x, y) | x^2 + y^2 = 4n^2 - 12n + 10, -1 \leqslant x, y \leqslant 2n-3, x + y \geqslant 2n - 4\}|.$$

The solutions $(\pm 1, 2n-3)$ and $(2n-3, \pm 1)$ lead to solutions $(1,1)$, $(1,2)$, $(n, n-1)$ and (n, n). It follows that

$$(3) \qquad S(n) = 2 + |\{(x, y) | x^2 + y^2 = 4n^2 - 12n + 10, x \geqslant 1, y \geqslant 1\}|.$$

According to Satz 161 and Satz 162 of E. Landau, *Elementare Zahlentheorie*, Chelsea, 1950, we obtain

$$(4) \qquad S(n) = 2 + \sum_{d^2 | (4n^2 - 12n + 10)} V\left((4n^2 - 12n + 10)/d^2\right)$$

where $V(m)$ is defined by

$$(5) \qquad V(m) = \begin{cases} 0 & \text{if 4 divides } m \text{ or if a prime } p,\ p \equiv 3 \pmod 4 \text{ divides } m, \\ 2^s & \text{if } m \not\equiv 0 \pmod 4 \text{ and no prime } p,\ p \equiv 3 \pmod 4, \\ & \text{divides } m \text{ and } s \text{ is the number of different odd divisors of } m. \end{cases}$$

Clearly we can replace (4) by

$$S(n) = 2 + \sum_{d^2|(2n^2-6n+5)} V((2n^2 - 6n + 5)/d^2). \tag{6}$$

(b) Since $2(2n^2 - 6n + 5) = (2n - 3)^2 + 1$ and since -1 is not a quadratic residue modulo a prime p when $p \equiv 3 \pmod 4$, it follows that

$$2n^2 - 6n + 5 = \prod_{i \in I} p_i^{e_i}, \quad e_i \geqslant 1, \tag{7}$$

where the p_i's are different primes congruent to 1 (mod 4). If $e_i \geqslant 2$ for some $i \in I$, then $d = 1$ and $d = p_i$ already lead to a value of $S(n)$ greater than 4 by (5). If $e_i = 1$ for all $i \in I$, then $S(n) = 2 + 2^{|I|} > 4$ if $|I| \geqslant 2$. It follows that $S(n) = 4$ iff $2n^2 - 6n + 5$ is a prime number.

(c) An odd term in (6) occurs by (5) iff $(2n^2 - 6n + 5) = d^2$. This condition can be rewritten as

$$(2n - 3)^2 - 2d^2 = -1, \tag{8}$$

and it follows from Theorem 8-6 in W. J. LeVeque, *Topics in Number Theory*, vol. I, Addison-Wesley 1956, that $S(n)$ is odd iff for some $m \geqslant 1$

$$2n - 3 = 1 + \binom{2m+1}{2}2 + \binom{2m+1}{4}2^2 + \cdots + \binom{2m+1}{2m}2^m. \tag{9}$$

Since the right-hand side in (9) equals

$$\frac{(1+\sqrt{2})^{2m+1} + (1-\sqrt{2})^{2m+1}}{2} = \langle (1+\sqrt{2})^{2m+1}/2 \rangle,$$

where $\langle x \rangle$ denotes the nearest integer to x, we have $S(n)$ is odd iff

$$n = \left(3 + \langle (1+\sqrt{2})^{2m+1}/2 \rangle\right)/2.$$

Parts (a), (b) also solved by J. Suck (Germany), and M. Vowe (Switzerland).

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor G. L. Alexanderson, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by April 30, 1984. The solver's full post-office address should be on each sheet.

6445. *Proposed by Richard Askey, University of Wisconsin.*

Show that Pfaff's transformation of the hypergeometric function

$${}_2F_1\left(\begin{matrix} a, b \\ c \end{matrix}; x\right) = (1 - x)^{-a}\,{}_2F_1\left(\begin{matrix} a, c-b \\ c \end{matrix}; \frac{x}{x-1}\right)$$

implies Landen's transformation of the dilogarithm

$$\mathrm{Li}_2 x := \sum_{n=1}^{\infty} \frac{x^n}{n^2} = -\frac{1}{2}[\log(1 - x)]^2 - \mathrm{Li}_2(x/(x - 1)).$$

6446. *Proposed by Paul S. Bruckman, Carmichael, CA.*

Define the generalized totient function ϕ_r as follows:

$$\phi_r(n) = \sum_{\substack{k=1 \\ (k,n)=1}}^{n} k^r, \quad r = 0, 1, 2, \ldots,$$

and its (Dirichlet) generating function

$$f_r(s) = \sum_{n=1}^{\infty} \phi_r(n)/n^s,$$

defined for an appropriate domain of s. Show that

$$f_r(s) = \frac{1}{(r+1)\zeta(s-r)} \sum_{k=1}^{r+1} \binom{r+1}{k} |B_{r+1-k}| \zeta(s-k), \quad s > r+2,$$

where the B_j's are Bernoulli numbers and ζ is the Riemann Zeta function.

6447. *Proposed by P. Erdős, Hungarian Academy of Sciences, and J. L. Selfridge, Mathematical Reviews.*

For fixed $k \geqslant 2$, if $n \geqslant 2k$, show that there is at least one i, $0 \leqslant i \leqslant k-1$, such that $n-i$ does not divide $\binom{n}{k}$. On the other hand, there is an $n_k \geqslant 2k$ for which $n-i$ divides $\binom{n}{k}$ for all but one i in the range $0 \leqslant i \leqslant k-1$. Estimate the smallest such n_k as well as you can.

SOLUTIONS OF ADVANCED PROBLEMS

The Expected Value of the Largest Part of a Partition

6386 [1982, 338]. *Proposed by I. P. Goulden and L. B. Richmond, University of Waterloo.*

Assuming each m-part ordered partition of n to have probability $1/\binom{n-1}{m-1}$, let the expected value of the largest part in such a partition be E_n. Show that, for fixed m, $\lim_{n\to\infty} E_n/n = S_m/m$, where $S_m = \sum_1^m 1/i$ is the mth partial harmonic sum.

Solution by Lajos Takács, Case Western Reserve University. If $n = \delta_1 + \delta_2 + \cdots + \delta_m$ is an ordered random partition of n into the sum of m positive integers, then $\delta_1, \delta_2, \ldots, \delta_m$ are interchangeable random variables and

$$P\{\delta_1 > k, \delta_2 > k, \ldots, \delta_j > k\} = \binom{n-jk-1}{m-1} \Big/ \binom{n-1}{m-1}$$

for $0 \leqslant k \leqslant (n-m)/j$ and $j = 1, 2, \ldots, m$, since the number of representations of n as a sum of m positive integers in which j given terms are greater than k is $\binom{n-jk-1}{m-1}$. If $\rho_m(n) = \max_{1\leqslant i\leqslant m}\delta_i$, then by the method of inclusion and exclusion we obtain that

$$P\{\rho_m(n) > k\} = \sum_{1\leqslant j\leqslant (n-m)/k} (-1)^{j-1}\binom{m}{j}\binom{n-1-jk}{m-1} \Big/ \binom{n-1}{m-1}$$

for $k = 1, 2, \ldots, n$.

For any positive integer j we have

$$\sum_{0\leqslant k\leqslant (n-m)/j} \binom{n-1-jk}{m-1} = \frac{1}{j}\binom{n}{m} + o(n^m)$$

as $n \to \infty$. (See [1].) Consequently,

$$E_n = E\{\rho_m(n)\} = \sum_{k=0}^{n} P\{\rho_m(n) > k\} = \frac{n}{m} \sum_{j=1}^{m} (-1)^{j-1}\binom{m}{j}\frac{1}{j} + o(n)$$

as $n \to \infty$. Since

$$\sum_{j=1}^{m} (-1)^{j-1}\binom{m}{j}\frac{1}{j} = \int_0^1 \frac{(1-x)^m - 1}{x}\,dx = \int_0^1 \frac{1-y^m}{1-y}\,dy$$
$$= \int_0^1 (1 + y + \cdots + y^{m-1})\,dy = \sum_{j=1}^{m}\frac{1}{j},$$

it follows that

$$\lim_{n\to\infty}\frac{E_n}{n}=\frac{1}{m}\sum_{j=1}^{m}\frac{1}{j}$$

for $m = 1,2,\ldots$.

Reference

1. L. Takács, A sum of binomial coefficients, Mathematics of Computation, 32 (1978) 1271–1273.

Also solved by Victor Hernandez (Spain), L. E. Mattics, Roger S. Pinkham, Helmut Prodinger (Austria), and the proposers.

The Sum of a Series

6387 [1982, 338]. *Proposed by Jeff Alden, Vector Research Inc., Ann Arbor, Michigan.*

Let a, y satisfy $-e^{-1} < a < e^{-1}$, $y = e^{ay}$. Prove that

$$\sum_{0}^{\infty}\frac{n^n a^n}{n!}=\frac{1}{(1-ay)}.$$

Solution by Mark A. Pinsky, Northwestern University. The additional condition $|ay| < 1$ is required to ensure that the correct solution of $y = e^{ay}$ is chosen. The result then holds for complex a with $|a| < e^{-1}$.

We begin with the identity

$$\sum_{n=0}^{\infty}\frac{n^n a^n}{n!}=\frac{1}{2\pi i}\int_{|z|=1}\frac{dz}{z-ae^z}.$$

This is proved by writing

$$\frac{n^n}{n!}=\frac{1}{2\pi i}\int_{|z|=1}\frac{e^{nz}}{z^{n+1}}dz,$$

multiplying by a^n and integrating the resulting uniformly convergent series term-by-term. By Rouché's theorem, the function $z - ae^z$ has exactly one zero z_1 in the unit disc $|z| < 1$. By the residue theorem the integral in question is then equal to $1/(1 - ae^{z_1})$, which yields the required result if we make the identification $z_1 = ay$.

Also solved by the proposer and 47 others, a number of whom pointed out that the result is known and referenced the solution to problem 214 on page 348 in *Problems and Theorems in Analysis* I, by G. Pólya and G. Szegö, Springer-Verlag, 1972.

On Rolling a Cube and a Tetrahedron

6388 (1982, 338]. *Proposed by Nicholas Wheeler and Howard Straubing, Reed College, Portland, Oregon.*

A regular tetrahedron R sits on a unit triangle T on a plane tiled with triangles congruent to T. A move consists in rotating R about an edge in contact with the plane. After several moves, R sits on T again. Have the vertices of R been permuted in space? What if R is a cube and the tiling is by squares?

Solution by Jerrold W. Grossman, Oakland University, Rochester, Michigan. The answer is no

for the tetrahedron and maybe for the cube. The latter is easy to see, since rotation north, then east, then south, then west, returns the cube to its starting position with the vertices permuted. To prove the former, note that the plane can be coordinatized so that the positive y-axis makes a 60° angle with the positive x-axis. Then the vertices of the triangular tiling are represented by pairs of integers (i, j). Reducing the coordinates modulo 2 we obtain a 4-coloring of the vertices of the partition. It is clear that if tetrahedron R, whose vertices are labelled by the same four "colors," is placed on the tiling so that colors match, then the matching is preserved by a rotation of R about an edge in contact with the plane.

Also solved by Miroslav D. Ašić (Yugoslavia), David A. Barrington, Benny N. Cheng, the Chico Problem Group, Hans E. Debrunner (Switzerland), Clayton W. Dodge, Pravin Kumar (India), Jeff Loveland, John R. Silvester (England), F. B. Strauss, David M. Wells, and the proposers.

A number of solvers showed that the cube can return to 12 of its 24 possible configurations. Hans E. Debrunner proved that a regular icosahedron in place of the tetrahedron can return to any of its 60 configurations, and a regular octahedron to 4 of its 24 configurations.

On a Rayleigh Quotient

6389 [1982, 339]. *Proposed by Alfonso Castro B., Centro Estudios Avanzados, Mexico, D. F.*

Let Ω denote a generic bounded region in real n-space $\mathbb{R}^n$. Let $H^1(\Omega)$ be the Sobolev space of square integrable functions in Ω having generalized first-order partial derivatives in $L^2(\Omega)$. (See R. Adams, *Sobolev spaces*, Academic Press, 1975.) With integrals over Ω, define

$$J(u) = \left[\sum_{i=1}^{n} \int (\partial u/\partial s_i)^2\right] \Big/ \int u^2.$$

Find a bounded region Ω such that the infimum (over the nonzero functions u in $H^1(\Omega)$ with $\int u = 0$) is 0.

Solution by the proposer. Let $c = \sum_{k=2}^{\infty} 1/k^2$, $d_n = \sum_{k=2}^{n} 1/k^2$, $(n = 2,3,\dots)$ and $d_0 = 0$. Let $\Omega = \{(x, y) \in \mathbb{R}^2;\ -1 < x < 0$ and $0 < y < 1$, or $d_{2n-1} \leqslant x \leqslant d_{2n}$ and $0 < y < 1/(2n+1)^4$ $(n = 2,3,\dots)$, or $d_{2n} < x < d_{2n+1}$ and $0 < y < 1$ $(n = 1,2,\dots)$, or $0 \leqslant x \leqslant 1/4$ and $0 < y < 1/81\}$.

Let $f_n\colon [-1, c] \to \mathbb{R}$ be defined by

$$f_n(x) = \begin{cases} \varepsilon_n & \text{if } x < d_{2n-1} \\ (1-\varepsilon_n)(2n)^2(x - d_{2n-1}) + \varepsilon_n & \text{if } d_{2n-1} \leqslant x < d_{2n} \\ 1 & \text{if } x \geqslant d_{2n} \end{cases}$$

where $-c < \varepsilon_n < 0$ is such that if $u_n(x, y) = f_n(x)$, then

$$\int_\Omega u_n(x, y)\, dx\, dy = 0. \tag{1}$$

Clearly $u_n \in H^1(\Omega)$ and

$$\int_\Omega u_n^2(x, y)\, dx\, dy \geqslant \sum_{k=n}^{\infty} 1/(2k+1)^2. \tag{2}$$

On the other hand

$$\int_\Omega (\partial u_n/\partial x)^2 + \int_\Omega (\partial u_n/\partial y)^2 = (1-\varepsilon_n)^2(2n)^2/(2n+1)^4. \tag{3}$$

From (2) we see that

$$\int_\Omega u_n^2 \geqslant \sum_{k=n}^{2n} 1/(2k+1)^2 \geqslant n/(4n+1)^2. \tag{4}$$

Combining (3) and (4) we see that $J(u_n) \to 0$ as $n \to \infty$, which proves that

$$0 = \inf\left\{ J(u);\ u \in H^1(\Omega), u \neq 0 \text{ and } \int_\Omega u = 0 \right\}.$$

Also solved by Alfonso Villani (Italy).

A Determinant

6390* [1982, 339]. *Proposed by P. Venzi and J. Aguadé, ETH, Zürich, Switzerland.*

Let $c_1, \ldots, c_n$ be real numbers, with $c_i \neq c_j$ if $i \neq j$. Consider the $n \times n$-matrix (a_{ij}) defined by $a_{ij} = c_j c_i/(c_j - c_i)$ for $i < j$, $a_{ji} = -a_{ij}$ for $i \neq j$ and $\Sigma_{j=1}^n a_{ij} = -c_i$, $i = 1, \ldots, n$. Prove that $\det(a_{ij}) = (-1)^n n! c_1 \cdots c_n$.

Solution by E. G. Straus, University of California at Los Angeles. We can factor out a factor c_i from the ith row to get

$$\det(a_{ij}) = c_1 c_2 \cdots c_n \det(b_{ij})$$

where

$$b_{ij} = \begin{cases} c_j/(c_j - c_i) & i \neq j \\ -1 - \sum\limits_{k \neq j} c_k/(c_k - c_j) & i = j. \end{cases}$$

We first prove that $\det(b_{ij})$ is a constant independent of the values of $c_1, \ldots, c_n$. Since it is clearly a homogeneous rational function of degree 0 in these variables, it suffices to prove that the denominator, which is a divisor of $\Pi_{1 \leqslant i < j \leqslant n}(c_i - c_j)^2$, is of degree 0.

By symmetry it suffices to prove that the factor $c_1 - c_2$ does not occur in the denominator. Since it occurs only in the denominators of $b_{11}, b_{12}, b_{21}, b_{22}$, we modify $\det(b_{ij})$ by first adding columns $2, 3, \ldots, n$ to the first column and then subtracting the first row from the second row. This yields a matrix (b'_{ij}) with

$$b'_{11} = -1,\ b'_{1j} = b_{1j} \text{ for } j > 2;$$

$$b'_{21} = 0,\ b'_{22} = -2 + \frac{c_3}{c_2 - c_3} + \cdots + \frac{c_n}{c_2 - c_n},$$

$$b'_{2j} = \frac{c_j}{c_j - c_2} - \frac{c_j}{c_j - c_1} = \frac{c_j(c_2 - c_1)}{(c_j - c_1)(c_j - c_2)} \quad \text{for } j > 2;$$

$$b'_{i1} = -1 \quad \text{for } i > 2;\ b'_{ij} = b_{ij} \quad \text{for } i > 1, j > 2.$$

Thus the only b'_{ij} with denominator having $c_1 - c_2$ as a factor is b'_{12}, and in $\det(b'_{ij})$ the cofactor of b'_{12} is a determinant with first row

$$(c_1 - c_2)\left(0, \frac{c_3}{(c_3 - c_1)(c_3 - c_2)}, \ldots, \frac{c_n}{(c_n - c_1)(c_n - c_2)}\right),$$

which is divisible by $c_1 - c_2$.

Having established that $\det(b_{ij})$ is a constant, it is easy to find its value. Set $d_{ij} = c_j/(c_j - c_i)$ if $i < j$. Then

$$b_{ij} = \begin{cases} d_{ij} & i < j \\ 1 - d_{ji} & i > j \\ -i + d_{1i} + d_{2i} + \cdots + d_{i-1,i} - d_{i,i+1} - \cdots - d_{in} & i = j. \end{cases}$$

Successively letting $c_1 \to \infty$, then $c_2 \to \infty$, etc., we find that every $d_{ij} \to 0$, and hence

$$\det(b_{ij}) = \begin{vmatrix} -1 & & & & & \\ 1 & -2 & & & 0 & \\ 1 & 1 & -3 & & & \\ & & \vdots & & & \\ 1 & 1 & 1 & \cdots & 1 & -n \end{vmatrix} = (-1)^n n!.$$

Also solved by Mihàly Bencze (Romania), C. S. Karuppan Chetty (India), Clark Givens, C. L. Mallows, and Y. S. Sathe & R. B. Bapat (India).

Prawn's Moves

6391 [1982, 339]. *Proposed by Louis W. Shapiro, Howard University, Washington, D.C.*

Let a prawn be a new chess piece that moves like a rook in one direction, and like a pawn in the perpendicular direction. A prawn starts in the lower left-hand corner of an $n \times n$ board. On each move, it can progress either any number of squares straight up (remaining on the board) or one square to the right.

(a) In how many ways can the prawn get to the mth diagonal (defined by $x + y = m$ on the standard lattice)? This sequence starts $1, 2, 5, 13, \ldots$.

(b) In how many ways can the prawn get to the opposite corner? This sequence starts $1, 2, 9, 44, \ldots$.

Solution by O. P. Lossers, Department of Mathematics and Computing Science, Eindhoven University of Technology, Eindhoven, The Netherlands. (a) The answer α_m equals f_{2m}, where $f_1, f_2, \ldots$ are the Fibonacci numbers given by $f_1 = 0$, $f_2 = 1$, $f_i = f_{i-1} + f_{i-2}$. Indeed, depending on whether the last move was of the pawn or of the rook type, the mth diagonal can be reached in α_{m-1} or $\alpha_1 + \alpha_2 + \cdots + \alpha_{m-1}$ ways respectively. Hence

$$\alpha_m = 2\alpha_{m-1} + \sum_{i=1}^{m-2} \alpha_i,$$

and consequently

$$\alpha_m = 3\alpha_{m-1} - \alpha_{m-2}.$$

Since $\alpha_1 = f_2 = 1$, $\alpha_2 = f_4 = 2$, it follows that α_m and f_{2m} satisfy the same recurrence relation and hence that $\alpha_m = f_{2m}$.

(b) To reach the opposite corner the prawn makes n pawn moves and k rook moves. These can be ordered in $\binom{n+k}{k}$ different ways. The k rook moves must add up to n steps upwards, and the number of possibilities for this is $\binom{n-1}{k-1}$, the number of partitions of n into k parts. It follows that the prawn can get to the opposite corner in

$$\sum_{k=1}^{n} \binom{n+k}{k}\binom{n-1}{k-1} =: \beta_n$$

different ways. From E. T. Whittaker and G. N. Watson, *Modern Analysis*, Cambridge, 1920, page 311, we quote the following expression for the Legendre polynomial $P_n(x)$:

$$P_n(x) = \sum_{k=0}^{n} \binom{n}{k}\binom{n+k}{k}\left(\frac{x-1}{2}\right)^k.$$

Differentiation yields

$$P_n'(x) = \frac{n}{2} \sum_{k=1}^{n} \binom{n-1}{k-1} \binom{n+k}{k} \left(\frac{x-1}{2}\right)^{k-1},$$

and hence

$$\beta_n = \frac{2}{n} P_n'(3).$$

Using a well-known recurrence relation for Legendre polynomials, we obtain the following recurrence relation for β_n:

$$\beta_0 = 1, \beta_1 = 2, \beta_{n+1} = \frac{3(2n+1)}{n+1}\beta_n - \frac{n-1}{n}\beta_{n-1}.$$

Also solved by Anthony E. Barkauskas, J. Binz (Switzerland), Roger Cuculiere (France), Roger B. Eggleton (Australia), Milton P. Eisner, C. Georghiou (Greece), Ira Gessel, Jerrold W. Grossman, J. Stephen Montague, Robert E. Shafer, Lajos Takács, John Ward, David M. Wells, and the proposer.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

A Number for Your Thoughts: Facts and Speculations About Numbers from Euclid to the Latest Computers. By Stephen P. Richards. S. P. Richards, New Providence, New Jersey, 1982. 207 pp. $7.95 (paperback).

UNDERWOOD DUDLEY

Department of Mathematics, DePauw University, Greencastle, IN 46135

Go to a mall. Go to a bookstore—a Waldenbooks, say. Go to the science section. What will you find? Four three-foot shelves: astronomy gets one shelf; weather, physics, and general science gets one shelf; biology gets one shelf; miscellaneous science gets one shelf. John McPhee is there, with *The Curve of Binding Energy*. Lewis Thomas is there. Isaac Asimov is there. *The Soul of a New Machine* is there. Is there any mathematics? One book and one only: Hofstadter's *Gödel, Escher, Bach*.

Go to another mall. Go to a B. Dalton's bookstore this time. The management is more ambitious than Waldenbooks', since there are at least five times as many titles on the shelves. General science is well represented, with books by Carl Sagan, Isaac Asimov, Jeremy Bernstein, and many others. There is Nigel Calder in the Physics section. There is the *Field Guide to the Atmosphere* in Earth Sciences, along with a misfiled *Catastrophe Theory*. There is even a lot of mathematics. But what kind of mathematics? *The Statistics Problem Solver*, *Advanced Engineering Mathematics*, *Forgotten Algebra* (*A Refresher Course*), only one copy of Courant and Robbins' *What is Mathematics*? and just two copies of Davis and Hersh's *The Mathematical Experience*. Most of the books are texts—sixty-one out of one hundred and one. There are seven reference works, four histories, five different works on mathematics anxiety. The books are almost all textbooks. The space devoted to astronomy is two and one-half times as large as that devoted to mathematics, and hardly any of the books are texts. Computer science gets twelve times as much space. Nature and the biological sciences are shelved separately, and their space is many times as great as the space given to mathematics.

Go to a public library. Go to the science section and see that the books on mathematics take up about one-twentieth of the total space for science. And what books they are! Three works on the slide rule, *Arithmetic for the Practical Man*, *Introduction to Plane Trigonometry*. Old books, undisturbed for years, the dust thick on them. I suspect that most of them were gifts to the library from people discarding books no longer of any use. Why would anyone want to keep a mathematics book?

Go to *Books in Print*, the 1982-83 edition. How many of the tens of thousands of books listed are classified under "Mathematics—Popular Works"? One hundred, would you think? Fifty? The number is eight. Eight only, and 37.5% of those were published in the Union of Soviet Socialist Republics. One of the remaining five is *Technical Shop Mechanics*, popular only in a restricted group. You can maintain, probably rightly, that the classifiers of the J. J. Bowker Company are not expert in classifying mathematics, but the fact that there are so few titles classified under Popular Mathematics shows that such works are rare and strange.

Why are there so few popular mathematics works? Is it that, as G. H. Hardy wrote, "Exposition, criticism, appreciation, is work for second-rate minds" so no one reads about mathematics because the books are of such low quality? I think not. Some first-rate minds have produced popular works in mathematics and, quality aside, many books in other sciences succeed. Is it because publishers are today almost all part of some conglomerate, so they no longer want to publish books that have no chance of selling a huge number of copies? Perhaps, but were many popular mathematics books being read back in the golden age of publishing, whenever that was? Aside from the success of a *Mathematics for the Million*, or a *World of Mathematics*, I think there was the same low level of readership as there is now. Is it that there is not enough advertising? Is it the fault of television? Air pollution? What can it be?

The answer is simple. People do not want to read about mathematics. They will not buy books about mathematics, no matter how well-written they are, no matter how many are published, no matter how hard they are advertised. They would rather read about other things. Quarks instead of primes. Black holes instead of Fermat numbers. Relativity instead of pi.

Why the difference? Of course, books on physics and astronomy can have pictures more attractive than portraits of Gauss and Euler, but I think it is no coincidence that the number of books published and read in mathematics, physics, and astronomy is inversely proportional to the number of people who studied each subject in school. It is our schools which have succeeded in making mathematics so distasteful and unattractive that the vast majority of the population wants to have nothing to do with the subject after school. If the study of mathematics was not inflicted on everyone, we might see almost as many popular works in mathematics as we do in physics.

It is curious that society insists that every citizen be taught to find solution sets of quadratic equations, not to confuse a number with a numeral, and to distinguish unary minus signs from binary ones when there is really no need for everyone to know these things. Mathematics is useful for some people and necessary for society, but so are the principles of sanitary engineering. It is curious because there is no great shortage of mathematical talent—if the financial rewards for mathematicians were as large as they are for, say, stockbrokers, graduate schools of mathematics would be turning away thousands of qualified applicants. It is curious because it is easy to spot mathematical talent, the utter lack of it is almost as easy to see, and no amount of schooling will change one or the other. Why indeed do we persist in continuing to force millions of suffering children to do something for which they have neither the talent nor the inclination?

The present situation is more than curious, even pernicious, because it creates dislike for mathematics and anxiety about it. Dislike creates bad feelings about mathematics and by extension about the entire life of the mind and anxiety creates bad feelings about oneself. How

much better to have a happier anxiety-free population! (That mathematics anxiety is not misplaced is shown by the record of professional anxiety-reducers. I have read no success stories of people who were once intimidated by decimals who are now proving theorems. The best that seems to happen is that some people are able to struggle through a semester of calculus with a grade of C.) To reduce anxiety is a waste of resources. To create dislike is a waste of resources. Waste is bad.

There is no hope that mathematics books will ever be popular, and *A Number for Your Thoughts* illustrates this. Its author loves numbers. He is crazy about them, and he wants you to be crazy about them too. Numbers are fascinating! They are at least as interesting as quarks, at least as rich as black holes, and at least as easy to understand as relativity. But the author has published his book himself. He had no publisher to tell him that references and an index are necessary. He had no editor to eliminate the embarrassing stylistic lapses and correct the occasional errors of fact. So his book will fail. In a more perfect world, a revision of it would have succeeded. But in this world, popular works in mathematics are doomed to be unpopular.

Emmy Noether, 1882–1935. By Auguste Dick. Translated by H. I. Blocher. Birkhäuser, Boston, MA, 1981. xiv + 193 pp.

Emmy Noether, A Tribute to Her Life and Work. Edited by James W. Brewer and Martha K. Smith. Marcel Dekker, New York, 1981.

Irving Kaplansky
Department of Mathematics, University of Chicago, Chicago IL 60637

Emmy Noether was born on March 23, 1882. Her numerous admirers have found much to cheer during the centenary year of 1982. In addition to the two books I am reviewing, Springer will shortly publish her collected works, edited by Jacobson. In March of 1982 a symposium in her honor was held at Bryn Mawr, and there will be a volume of proceedings. So, in all, we have three books as a 100th birthday salute, with a fourth coming before long.

I am a member of the steadily shrinking circle of people who have a personal recollection of Emmy Noether (when I was a freshman at Toronto she gave a colloquium talk). But I did not know at first hand the German mathematical scene between the first World War and 1933. I picture a "golden age" of ferment, excitement, and great progress. At the center was Göttingen, and algebra at Göttingen meant Emmy Noether. In 1933 her stature as a superb mathematician was firmly established. Yet, at the age of 51, her rank was "ausserordentlicher Professor" (associate professor, more or less). Even this had been achieved only by a concerted battle against prejudice.

And then the storm broke. On April 2, 1933, her right to teach at Göttingen was withdrawn. The letter from the Prussian Ministry of Science, Art, and Public Education cited as its authority the new statutes of the National Socialist regime (which, for a time at least, gave lip service to the German tradition of orderly procedure).

"It shall not be forgotten what America did during these last two stressful years for Emmy Noether and for German science in general." These are the words of Hermann Weyl. Perhaps they are overgenerous. At least initially, the positions found for the illustrious immigrants (as Laura Fermi called them) were modest. Still, it was the bottom of the depression, and action had to be taken quickly. I think that the North American mathematical community can be reasonably proud of what was achieved. The reward was ample; there is no doubt that this infusion played a major role in transferring mathematical leadership from Germany to the U.S.A. Today another immigra-

tion is taking place; it is smaller but again illustrious. Our colleagues in the Soviet Union should ponder carefully the lessons of the past.

In Emmy Noether's case the transition was from Göttingen to Bryn Mawr till her untimely death on April 14, 1935. (An interesting sidelight is the fact that Edward R. Murrow, then with the Institute of International Education, played a role in the negotiations that brought her to Bryn Mawr.)

Dick's biography was published in German in 1970. It is likely to remain the standard work. The translation into English incorporates some minor changes. It will make the story of her life more widely available and I accordingly welcome it.

I recommend to the reader the review by B. H. Neumann in the Bulletin of the London Mathematical Society (vol. 14, pp. 155–6). Some of his comments have my endorsement. I cannot resist mentioning that the typo he spotted (a misspelling of C. C. Faith on page 98) is not the only one: Herbrand is misspelled in the caption facing page 75.

Completing the book are the obituaries by van der Waerden, Weyl, and Alexandrov, a bibliography, and a list of her Ph.D.'s.

The volume edited by Brewer and Smith consists almost entirely of new material. Clark Kimberling, who had much to do with initiating the book, begins with a much expanded version of his 1972 MONTHLY biography. But I urge readers nevertheless to look up the previous article for the fascinating final section on the recovery of letters of Cantor, Dedekind, Frobenius, and Weber after a 33-year sojourn in a Philadelphia law office.

The book continues with incisive reminiscences by Saunders Mac Lane of Göttingen in 1931–3, together with a letter he wrote to his mother on December 8, 1931. A reader might enjoy supplementing this with the description of the writing of his Göttingen thesis that appears in his *Selected Papers* (Springer, 1979).

Olga Taussky had a unique opportunity to observe Emmy Noether: in Göttingen in 1931–2 and in Bryn Mawr in 1934–5. In my opinion her 20 pages of reminiscences bring Emmy Noether to life more than anything else in either book.

In the next two chapters we find the van der Waerden and Alexandrov obituaries. These duplicate the corresponding portions of the Dick book. But it is not complete duplication, for the reader can have the fun of comparing two different translations.

Available in English for the first time is her 1932 address to the Zürich International Congress.

Five mathematicians contribute articles analyzing various aspects of Noether's work. In alphabetic order they are Fröhlich, Gilmer, Lam, McShane, and Swan. These papers add up to a contribution to the mathematical literature of lasting importance, and I move a vote of thanks.

I feel that McShane's article deserves special mention. I dare say it may surprise some of my readers to learn that many physicists are quite familiar with the name of Emmy Noether. But to them it does not conjure up commutative rings with the ascending chain condition, or even the mathematician who taught us the power of abstract algebra. They are thinking of a theorem about problems in the calculus of variations for which there is a group of transformations. The relevant paper is entitled "Invariante Variationsprobleme" and it appeared in the 1918 Göttinger Nachrichten.

It was not through publications alone that her influence spread. Here are some of her Ph.D.'s: Grell, Levitzki, Deuring, Fitting, Witt, Tsen. Many others, such as Alexandrov, Artin, Herbrand, Krull, and van der Waerden, were deeply influenced. The circle of students and disciples that surrounded her in Göttingen was jokingly referred to as the "Noether boys." After the transplantation to Bryn Mawr, she worked with four "Noether girls." Three were postdoctoral fellows: Olga Taussky, Marie Weiss, and Grace Shover (Quinn). Ruth Stauffer (McKee) received a Bryn Mawr Ph.D. under her direction.

The spirit of Emmy Noether lives on not only in her mathematical works, but also in her many students and colleagues, and in their mathematical progeny as well.

The Theory of Spinors. By Elie Cartan. Dover, New York, 1981. (Unabridged republication of the complete English translation first published by Hermann of Paris in 1966.) x + 157 pp. $4.00.

ROBERT HERMANN
APSM, 53 Jordan Road, Brookline, MA 02146

Cartan is certainly one of the greatest and most original minds of mathematics, whose work on Lie groups, differential geometry, and the geometric theory of differential equations is at the foundation of much of what we do today. In my view, his place in mathematics is similar to that of the great turn-of-the-century masters in other areas of intellectual life. Just as Freud was influenced by the mechanistic world view of 19th century science, but used this background to create something new and revolutionary which has profoundly influenced 20th century thought, so Cartan built, on a foundation of the mathematics which was fashionable in the 1890's in Paris, Berlin and Gottingen, a mathematical edifice whose implications we are still investigating. His work was highly intuitive and geometric, but was also based on a formidable combination of original methods of calculation and analysis, ranging in mathematical expertise from algebra to topology. For example, he completed the work of Killing and Lie on the classification of simple Lie algebras. As Hawkins so convincingly demonstrates **[1]**, this required the mastery of the most advanced algebraic technique of the 1890's, a task at which Killing himself (who learned his algebra from Weierstrass!) had despaired. In the 1920's, when he was already in his 50's, he proved that the second homotopy group of a Lie group was zero, which was one of the first great general theorems about topology. As one can see in his Collected Works, he was a master of brutal calculations, and all of his work was based on an intimate knowledge of computational details and examples. In short, he was comparable to such great figures of mathematics as Gauss, Riemann, and Poincaré.

Thus it is unfortunate that his original work is so inaccessible to the wide spectrum of mathematicians and scientists who now make use of it. (Of course, it is accessible through the expositions by many others in the last twenty years.) This short book is a translation of one which was written in French in the 1930's, and is perhaps the most readable of his works. It was first published in 1966, when the work of Killing and Cartan on the classification of simple Lie groups was beginning to be applied in elementary particle physics. In terms of contemporary Lie group theory, it deals with the B and D series of simple Lie algebras and the Lie groups which go along with them, i.e., the orthogonal matrix groups over the real and complex numbers and their simply connected covering groups. Cartan starts off in complex Euclidean 3-space, describing "spinors" (which are elements of the vector space which carries the lowest-dimensional representation of the Lie algebra of 3×3 skew-symmetric real matrices) rather concretely and computationally, in a form which would satisfy the most down-to-earth physicist. He then goes on to generalize to the n-dimensional case in an explicit, but, perhaps inevitably sketchy, way. Any of the topics treated here would be useful reading for a physicist trying to understand contemporary Lie group theory or for a mathematics student trying to understand how linear algebra is applied in group theory. The last brief chapter sketches how the ideas can be combined with his "method of the moving frame" to describe how the Dirac equation of quantum mechanics can be formulated for an arbitrary Riemannian manifold. This material is of historical interest in physics, since it foreshadows the recent "gauge theory" developments. Cartan also takes a certain delight in pointing out in the Introduction that he discovered Spinors before Dirac, in 1913, as a by-product of his classification of representations of semi-simple Lie groups.

I hope I am not insulting the memory of my greatest hero to say that this book is a fraud! I don't believe that Cartan thought about the subject in the form in which it is presented here. Clearly, he is presenting a "vulgarization" of the general theory of semi-simple Lie algebras and groups, which he developed almost single-handedly (with the help of Hermann Weyl!) in the

period 1893-1930. Cartan was very much a fan of physics, and he clearly is trying to teach the physicists of his day some of his profound knowledge in a form which they might find more palatable. The recently published correspondence between Einstein and Cartan [2] is very illuminating about the habits of mind of these two great men, and even somewhat sad. They were like ships passing in the night: Cartan enthusiastically tried to communicate some of his great geometric ideas to Einstein, who was rather closed-minded and even condescending.

Finally, I would like to use this occasion to convey my thanks to Dover Publications. Their efforts to keep in print, at modest prices, the classics of science have been a great help in my own work. Many times I have picked up a Dover book at my local bookstore and discovered a jewel of which I was previously unaware. Even more important, it is one of the last links in today's world with certain precious scholarly traditions.

References

1. T. Hawkins, Wilhelm Killing and the structure of Lie algebras, Arch. Hist. Exact Sci., 26 (1982) 127–192.
2. E. Cartan and A. Einstein, Letters on Absolute Parallelism, Princeton University Press, 1979.

A First Course in Group Theory. By Cyril F. Gardiner. Springer-Verlag, New York, 1980. ix + 227 pp.

RONALD SOLOMON
Department of Mathematics, Ohio State University, Columbus, OH 43210

The label "group" attaches itself to quite a menagerie of beasts. An introductory text can easily attain encyclopedic proportions, looming like a Mount Everest for the intrepid. This text, however, is for the more timorous who prefer a pleasant summer's stroll through a few basic ideas and past several interesting examples. To motivate this brief excursion an ever-so-modest goal is proposed—the determination of all groups of order 1 through 15. In pursuit of this goal some important results are established, notably the structure theorem for finitely generated abelian groups and the Sylow theorems. Of course, the goal is clear and easy to state; it is, however, rather flimsy justification for the heavy artillery mobilized to attack it!

Does the world need such a modest text? Certainly there is nothing here which is not already treated in greater depth in Zassenhaus' classic text, as well as several successors. Usually a student is introduced to group theory at this level in a basic algebra course for which excellent texts such as Herstein's *Topics in Algebra* are available. What then is the audience for such a book? Perhaps there is a place for it in a freshman or sophomore level "short course," designed to give students a taste of noncalculus mathematics. Such courses have been proposed at Ohio State, but none has been offered, so I cannot speak from firsthand experience. Nonetheless, I believe such a course could be valuable and appealing to students. Accepting this as a raison d'être for the text, we can examine how it serves its audience.

The elementary approaches to the study of finite groups are of two types: the permutation methods and the "normal subgroup" methods. The former feature combinatorial arguments and tend to apply only to finite groups. The latter are often instances of much more general techniques of value in the study of all groups, rings, and modules. How do each of these approaches fare in this text?

Permutation arguments abound, but the role of permutation representations as a unifying idea in group theory is never made clear. On the one hand the standard lemmas on orbits and stabilizers are presented and used to prove Burnside's Counting Theorem with a (much too

trivial!) application to counting colored beads. On the other hand cosets and double cosets are introduced as tools for proving Lagrange's theorem and Sylow's theorems. But the two hands are never clasped! Nowhere is it mentioned that the cosets Hg of H in G are the H-orbits of the left regular permutation action on G or that the (H, K) double cosets are the K-orbits on $H \setminus G$. Such a unification of apparently disparate ideas would have lent beauty and clarity to the presentation, as well as considerably shortening some proofs.

The "normal subgroup" approach fares even worse. Although quotient groups are introduced early, some of the basic isomorphism theorems are deferred to the "epilogue." Automorphisms are defined only in an exercise, although they make a prior guest appearance in the definition of a semidirect product. What a pity! The study of automorphisms of cyclic groups provides a lovely arena for the introduction of some elementary number theory and the appreciation of its interface with group theory. Depending on the background of the students, some interrelations between group theory and linear algebra could also be explored.

Given this slighting of automorphisms, it is not surprising that semidirect products are banished immediately after their introduction. This is the one unforgivable failing of the text! Although the subject can become confusing to novitiates, even a brief effort would mobilize a powerful weapon for the construction and classification of groups. With this in hand the author could have quickly disposed of groups of order 1 to 15 or even consigned them to the exercises where they belong. This would have impressed students with the force of the techniques developed. Lacking this, he is forced to identify groups by their multiplication tables, a tedious and page-consuming process, hardly designed to convert the uninitiated to the power or beauty of group theory.

On the positive side, the highlight of the book is the unit on finite subgroups of $O(\mathbb{R}^3)$. The classification of these subgroups is a lovely elementary result of considerable interest both within mathematics and in the physical sciences; it is neglected by most basic texts. It is disappointing that this material is not given more space and prominence! Surely for an audience of beginners nothing could be more fun than to explore the notions of symmetry and group actions using models of the Platonic solids! As it is, the tetrahedral, octahedral and icosahedral groups are never identified as isomorphic to A_4, S_4 and A_5, respectively, although these identifications are well within the scope of the text. Indeed the first is trivial, the second is easy and the third may be inferred from a geometrical observation about the icosahedron, which the students could discover themselves. A second approach to the icosahedral group could use the action on pairs of antipodal vertices to identify it as a subgroup of S_6 of order 60. Then Sylow theory could be invoked to establish that any subgroup of S_6 of order 60 is isomorphic to A_5. This would serve as a nontrivial illustration of Sylow methods. With the concept of semidirect products in hand, another fun topic could be explored: Euclidean groups, tilings of $\mathbb{R}^2$ and Escher's graphics.

The concluding chapter or "epilogue" touches on the general structure problem for finite groups, subdivided into the extension problem and the finite simple group problem. Some measure of the difficulty of the latter problem is hinted at by reference to the length and depth of the Feit-Thompson theorem. The former problem, however, receives a misleading treatment, which gives no sense of its complexity and might indeed be construed as equating the extension problem (the building of a new group out of smaller ones) with the composition series problem (the demolition of a given group into its constituent bricks).

One final cavil concerns the exercises which are almost all rather pedestrian, but for which the author insists, nonetheless, on furnishing solutions. A bit more confidence in the ingenuity of students could have yielded a much meatier and more challenging volume.

I am convinced that an audience exists for such a book but there is no text at this level which I can recommend. In terms of style and enthusiasm to communicate, the book under review is exemplary. Unfortunately, its flaws, particularly the belaboring of multiplication tables and groups of small order, render it unacceptable as a course text.

LETTERS TO THE EDITOR

Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Editor:

I enjoyed reading the article entitled "A Technique for Integration by Parts" by Herbert E. Kasube (this MONTHLY, March 1983, pp. 210–211). The author gave the impression that his method for choosing u and dv will always work. Although this technique is usually helpful in choosing u and dv, I have concluded that sometimes his technique should not be used.

As an example, consider $\int \frac{xe^x dx}{(x+1)^2}$. Using Professor Kasube's technique, we would choose $u = \frac{x}{(x+1)^2}$ and $dv = e^x dx$. Thus $du = \frac{(-x+1)\,dx}{(x+1)^3}$ and $v = e^x$.
Therefore

$$\int \frac{xe^x dx}{(x+1)^2} = \frac{xe^x}{(x+1)^2} - \int \frac{(-x+1)e^x dx}{(x+1)^3}$$
$$= \frac{xe^x}{(x+1)^2} + \int \frac{xe^x dx}{(x+1)^3} - \int \frac{e^x dx}{(x+1)^3}.$$

If we continue using Kasube's technique on $\int \frac{xe^x dx}{(x+1)^3}$ and $\int \frac{e^x dx}{(x+1)^3}$, we will not progress to a solution.

However, the evaluation of $\int \frac{xe^x dx}{(x+1)^2}$ can be easily accomplished if we choose $u = xe^x$ and $dv = \frac{dx}{(x+1)^2}$.

David H. Brown
Department of Mathematics
Niagara County Community College
3111 Saunders Settlement Road
Sanborn, NY 14132

MISCELLANEA

117.

Others believe that if there is one word that expresses the spirit of the age, it is parameter, a mathematical term now widely misused so that nobody finds himself in the hateful position of having to say boundary or limit.

—Edwin Newman, *Strictly Speaking*, Warner Books, 1974, p. 17

MISCELLANEA

ERRATA AND ADDENDA

Referring to his joint paper with Charles Blair, "A universal entire function," which appeared in this MONTHLY, 5 (1983) 331–332, Lee A. Rubel writes as follows: "Robert Burckel has kindly drawn our attention to the paper by Gerald MacLane, 'Sequences of derivatives and normal families,' J. Analyse Math., 2 (1952) 72–87 [MR 14, p. 74] [Zbl. 49, p. 57]. Among other things, MacLane constructed 'a ubiquitous entire function' by the same method we use, to have the same 'universal' property. However, he also gets estimates on the rate of growth of the function, so that his proof is somewhat more complicated than ours."

U.S. Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U.S.C. 3685)

1A TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY	1B PUBLICATION NO: A M M Y A E	2 DATE OF FILING: Sept. 9, 1983
3 FREQUENCY OF ISSUE: Monthly except July and August	3A NO. OF ISSUES PUBLISHED ANNUALLY: Ten	3B. ANNUAL SUBSCRIPTION PRICE: Library $50.00, Indiv.Member $20.00

4 COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street, City County State and ZIP Code) (Not printers)*

1529 Eighteenth Street, N.W., Washington, D.C. 20036

5 COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OF GENERAL BUSINESS OFFICES OF THE PUBLISHER *(Not printer)*

1529 Eighteenth Street, N.W., Washington, D.C. 20036

6. FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*

The Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, D.C. 20036

EDITOR *(Name and Complete Mailing Address)*

P.R. Halmos, Indiana University, Dept. of Mathematics, Bloomington, IN 47405

MANAGING EDITOR *(Name and Complete Mailing Address)*

Dr. Raoul Hailpern, SUNY at Buffalo, Buffalo, NY 14214

7 OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

FULL NAME	COMPLETE MAILING ADDRESS
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W. Washington, D.C. 20036

8 KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

FULL NAME	COMPLETE MAILING ADDRESS
None	

9 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423.12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS
(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS
(If changed, publisher must submit explanation of change with this statement.)

10 EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A TOTAL NO. COPIES *(Net Press Run)*	20,699	18,349
B PAID CIRCULATION 1 Sales through dealers and carriers, street vendors and counter sales	10	10
2 Mail Subscription	17,626	15,933
C TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	17,636	15,943
D FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	90	90
E TOTAL DISTRIBUTION *(Sum of C and D)*	17,726	16,033
F COPIES NOT DISTRIBUTED 1 Office use, left over, unaccounted, spoiled after printing	2,943	2,316
2 Return from News Agents	none	none
G TOTAL *(Sum of E, F1 and 2 – should equal net press run shown in A)*	20,669	18,349

11 I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Alfred B. Willcox, Executive Director

PS Form 3526, July 1982 *(See instruction on reverse)*

Reston
Has Your Next Computer Textbook

DISCRETE MATHEMATICS FOR COMPUTER SCIENTISTS
Joe L. Mott, Abraham Kandel, Theodore P. Baker, all of Florida State University

Requires no specific mathematical background outside of the material ordinarily covered in most college algebra courses. Particularly useful for teachers who are mathematicians but not computer scientists by profession or training.

R1372-3 350 pp. 1983 $24.95

COMPUTABILITY
George J. Tourlakis, York University

An authoritative, comprehensive introduction to recursion theory. Offers a very good balance of formalism and intuition.

R0876-4 576 pp. Sept. 1983 $27.95

MATH COMPANION FOR COMPUTER SCIENCE
Zamir Bavel, University of Kansas

Widely adopted text that thoroughly discusses the math concepts and techniques most often encountered in computer science and presents them in an easy-to-use format.

R4300-1 384 pp. 1982 $21.95 cloth
R4299-5 384 pp. 1982 $16.95 paper

PROBLEM SOLVING FOR DISCRETE STRUCTURES
Zamir Bavel, Roxy Peck, Jan Bergandy, and Jerzy Grymala-Busse

An invaluable supplement that furnishes numerous exercises pertinent to the study of discrete structures. Complements the Math Companion for Computer Science chapter for chapter, but can be used for other texts.

R5618-5 256 pp. Dec. 1983 $12.95 paper

INTRODUCTION TO THE THEORY OF AUTOMATA
Zamir Bavel, University of Kansas

Covers both the algebraic and the combinatorial aspects of automata. Excellent treatment of sequential machines, sequential tasks and processes, and finite memory programs and devices.

R3271-6 650 pp. 1983 $27.95

MATHEMATICS FOR DATA PROCESSING, 2nd EDITION
Frank J. Clark

This best-selling text provides a broad background for students seeking an entry level position in data processing. A very solid integration of modern algebra, Boolean algebra, and programming techniques.

R4263-1 326 pp. 1983 $21.95

For adoption consideration, please write for examination copies on school letterhead to:

Martin Conner, Marketing Manager
Reston Publishing Company, Inc.
11480 Sunset Hills Road
Reston, VA 22090

Prices subject to change without notice.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.